Graph Partition Based on Dimensionless Similarity and Its Application to Fault Diagnosis

Bo Zheng[1, 2], Hong-Zhong Huang[1], Jie Zhou[1], and Yan-Feng Li[1]

[1]*Center for System Reliability and Safety, University of Electronic Science and Technology of China,*

Chengdu, Sichuan, 611731, China

b_zheng1@126.com
hzhuang@uestc.deu.cn
zhoujseven@gmail.com
Yanfengli@uestc.edu.cn

[2]*Civil Aviation Flight University of China*
Guanghan, Sichuan, 618307, China

ABSTRACT

To improve the efficiency of fault diagnosis, a novel granular computing algorithm is developed to reduce computational cost. It is realized by extracting and partitioning on the complete graphs, and in the process of graph generation, the graph partition based on dimensionless similarity (GPDS) method is proposed to overcome the influence of attributes with different dimensions. The algorithm is named graph partition based on dimensionless similarity. Moreover, similarity threshold determination method based on frequency distribution histogram is proposed to reduce the dependency on the experiences of experts. Meanwhile, a weighted relative error is proposed to measure quantitatively the distribution change of original data after being compressed. Finally, different characteristic data are applied to verify the theories. The experimental results indicate that the compressed training samples can maintain the classification accuracy. Furthermore, the elapsed time can be obviously reduced. Therefore, the GPDS method can be used in fault diagnosis properly

Keywords: Granular Computing; Graph Partition; Dimensionless Similarity; Weighted Relative Error; Fault Diagnosis

1. INTRODUCTION

With the development of modern industry, fault diagnosis technologies are widely used. Due to the rapid improvement of sensor technology and intelligence algorithms, the fault diagnosis based on data-driven has been widely used in solving particle large and complicated equipments (Zheng & Gao, 2015). Undoubtedly, numerous application cases of fault diagnosis field have indicated that it can be used to improve the efficiency of troubleshooting, shorten the maintenance period, reduce the maintenance costs and ensure production safety. However, many similar or repeated monitor data reflecting the same running state of equipments are recorded, which will obviously increase the sample size, and the computational efficiency of diagnosis algorithms will be reduced. So quick and accurate diagnosis has become a new research direction in fault diagnosis.

Granular computing (GrC) is new intelligent computing theory based on partition of problem concepts (Skowron & Stepaniuk, 2001). Granular computing can be used to reduce complex problems for obtaining the satisfactory approximate solutions at a lower computational cost. Meanwhile, it can be used to hide or reveal some details of the problems by changing the size of granulation. At present, there are three theories about granular computing, i.e. quotient space (Zhang & Zhang, 1992 and 2003), rough set (Z. Pawlak, 1991), and computing with words (Zadeh, 1996 and 1997). The detailed comparisons between aforementioned three theories have been discussed in (Y. Li, 2007; Zhang et al., 2004; Zhao & Yang, 2007). Moreover, some machine learning approaches, e.g. correlation calculation, decision tree, clustering algorithms and neural networks are used to generate granular computing model (Tang et al., 2005; Zhong et al., 2007; Anjum et al., 2009; Park et al., 2009), the theories analyze the realization process of granular computing from different angles. It is well known that granular computing opens up a new direction for intelligent computing.

The purpose of this research is to reduce the training samples and thus improve that the calculation efficiency. Considering the characteristics of monitor data, granular

computing can realize the data compression properly. But in practice, accuracy is a prerequisite for fault diagnosis, so the accuracy should not be changed after data compression. Compared with three main theories mentioned above, the granular is constructed in this paper depending on graph partition based on dimensionless similarity (GPDS) rather than equivalence class, or fuzziness of semantic expression. Thus, the granulation is described according to similarity, in essence, lower similarity means coarser granular, and higher similarity means finer granular. Therefore, the granular computing can be realized from a new angle.

The proposed granular computing method is designed to follow the principle of maintaining the original spatial distribution of data. In this condition, an evaluation index based on weighted relative error is designed to quantitatively measure the distribution change of original data after compression. Accordingly, although the number of training samples is reduced, the spatial distribution characteristic is similar to the original distribution, which is conducive to identify the fault data with relative lower computational cost.

The rest of the paper is organized as follows. In Section 2, dimensionless similarity is defined and the graph is generated by using monitoring data. In Section 3, the GPDS is introduced in details. Section 4 introduces the application of granular and weighted relative error. After that, the application problems have been used to verify the performance of GPDS. Finally, Section 5 draws conclusions.

2. SIMILARITY DEFINITION AND GRAPH GENERATION

2.1. Similarity Definition

Actually, many definitions of similarity have been used to measure the similar relationship of vectors, such as, Minkowski distance, Mahalanobis distance, Manhattan distance, Chebyshev distance, Pearson correlation coefficient, and cosine similarity [15-18]. These definitions measure the similarity according to different theories. Generally, Euclidean distance has been widely used, and several popular algorithms are computed by Euclidean distance (Neto, 2014; Zheng et al., 2015; Airteimoori & Kordrostami, 2010)

Euclidean distance is in the interval [0, $+\infty$), where the lower value means the more similarity, vice versa. In order to describe the granulation and divide the granular level conveniently, the similarity based on Euclidean distance is set in the interval (0, 1], "1" represents two vectors are identical; if the Euclidean distance of two vector is infinite, the similarity will be zero. In previous researches, Zhang (2000) and von Luxburg (2007) respectively proposed similar calculation approaches of exponential similarity. Huang (2007) proposed a calculation approach of inverse distance similarity. These similarities are in the interval (0, 1] which had been applied in practical computations.

Based on our work, we find that the similarities based on the traditional Euclidean distance have obvious defect when processing the vectors with different dimensions. Specifically, different dimensions in a vector mean the elements are in different numerical distribution ranges, which will easily lead to the similarity approach zeros. Undoubtedly, it has limited the application of similarity significantly. In this study, a dimensionless similarity is proposed to overcome the problems. If $\boldsymbol{X}$ denotes a samples set, $\boldsymbol{X}=[\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n]$, and $\boldsymbol{x}_i=[x_{i1}, x_{i2}, \cdots, x_{id}]$, $|\boldsymbol{x}_i|=d$, and d is the dimension of each sample.

Traditional similarities only consider the similar relation of two samples separately, and the dimensionless similarity sufficiently considers the influence of all samples on calculating Euclidean distance. Thus, the range of each attribute is introduced to eliminate dimension so that data have the same caliber. Accordingly, the dimensionless similarity between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is designed as follows:

$$\begin{cases} S(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{1}{\sqrt{2}} \cdot \sqrt{\sum_{k=1}^{d} \frac{\left(x_{ik} - x_{jk}\right)^2}{r_k}} \right) \\ r_k = \max(x_{1k}, x_{2k}, \cdots, x_{nk}) - \min(x_{1k}, x_{2k}, \cdots, x_{nk}) \end{cases} \quad (1)$$

where r_k represents the range of the k-th column. In Eq. (1), although $S(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is also in the interval (0, 1], the defect of similarity based on the traditional Euclidean distance is overcome.

2.2. Graph Generation

Graph partition is used to generate granular. In this theory, a triple can be used to describe the graph system ($\boldsymbol{GS}$), which can be defined as follows:

$$\boldsymbol{GS} = \left(\boldsymbol{X}, s, \boldsymbol{M} \right), \quad (2)$$

where $\boldsymbol{X}$ denotes the nonempty samples set formed by objects, it is called a universe of discourse, $\boldsymbol{X}=[\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n]$, $n=|\boldsymbol{X}|$; s is similarity threshold. Lower similarity generates coarser granular, and higher similarity generates finer granular. $\boldsymbol{M}$ is similarity matrix, the element of similarity matrix $e_{i,j} \in \{0,1\}$, it can be calculated by Eq.(3):

$$e_{i,j} = \begin{cases} S(\boldsymbol{x}_i, \boldsymbol{x}_j), S(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq s \\ 0, \text{otherwise} \end{cases} \quad (3)$$

Due to $\left\| \boldsymbol{x}_i - \boldsymbol{x}_j \right\|_2 = \left\| \boldsymbol{x}_j - \boldsymbol{x}_i \right\|_2$, and $\left\| \boldsymbol{x}_i - \boldsymbol{x}_i \right\|_2 = 0$, similarity matrix is expressed in Eq .(4):

$$\boldsymbol{M}_{n\times n}=\begin{bmatrix}0 & e_{12} & \cdots & e_{1n}\\ 0 & 0 & \cdots & e_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 0\end{bmatrix}, \tag{4}$$

Actually, similarity matrix is upper triangular matrix, and all diagonal elements are 0. In the similarity matrix, if $e_{ij}\neq 0$, a line will connect two samples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, and the value of e_{ij} is the weight of the line. Otherwise, there is not any connecting lines between two samples. According to the ***GS***, the graph ***G*** of objects can be generated, and ***G*** = (***X***, ***M***), actually, where ***X*** represents the set of points, and ***M*** represents the set of edges.

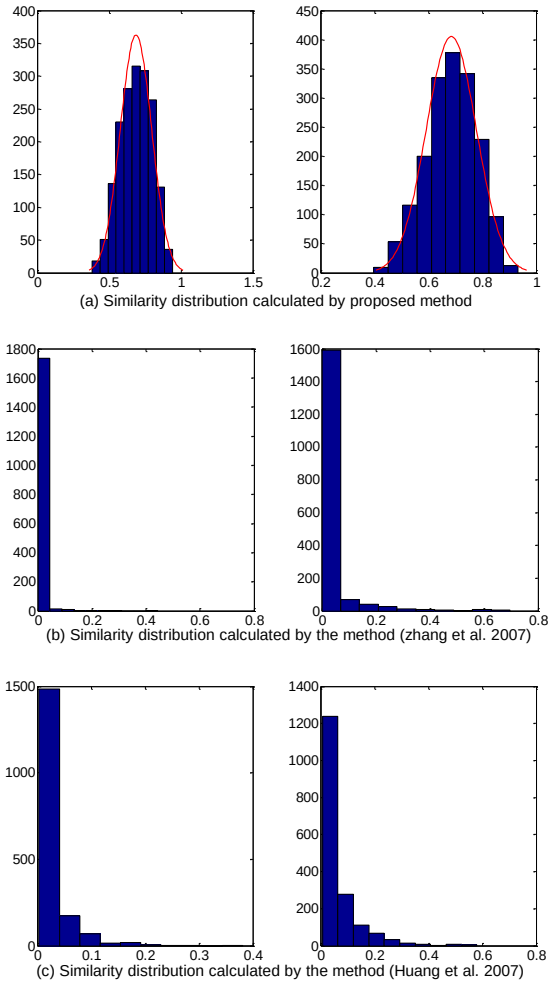


Figure 1. Comparison of similarity distribution calculated by different methods

A difficult and important issue is how to determine the similarity threshold *s* (Zhang et al., 2000; Luvburg, 2007; Huang et al., 2007). Generally, *s* can be given by trial-and-error method, or experts according to real problem. However, the method depending on the experts' experiences is inevitably subjective. In this paper, the numerical determination method based on the frequency distribution histogram is proposed, which can be used to illustrate frequency distribution of similarity matrix. In the proposed method, the interval (0, 1] is divided into 10 equal subintervals, and then the quantity of elements belonging to each subinterval can be counted. Normally, the distribution of all elements is similar to normal distribution. In practice, the similarity threshold *s* can be adjusted depending on the frequency distribution histogram.

The advantage of proposed method is that it can give the distribution of similarity matrix intuitively, which is conducive to reduce the dependency on the experiences of experts, meanwhile, the granular generation can be realized. Furthermore, the data of bearing outer and inner race fault (Zheng et al., 2016) will be used to verify the merit of dimensionless similarity by comparisons with the methods cited from Refs. (Zhang, 2000) and (Huang et al., 2007), the two kinds of fault respectively contain 60 samples, and the attributes of sample have different dimensions. The comparisons of similarity frequency distribution are shown in Figure1.

According to Figure 1, due to the influence of different dimensions, the similarity calculated by the methods from Refs. (Zhang, 2000) and (Huang et al., 2007) can hardly be used from the engineering standpoint, and the distributions are not norm distribution. The comparison results indicate that dimensionless similarity is rational and effective for engineering practice.

3. GRAPH PARTITION METHOD BASED ON SIMILARITY

In the graph theory, graphs containing no loops or multiple edges will be referred to as simple graphs, and the graph ***G*** generated by ***GS*** is a weighted simple graph. In the graph ***G***, the trivial subgraph $\boldsymbol{G}_{ts}$ having no connections with other points will be regarded as granular that only containing one point. Furthermore, except all trivial subgraphs, the subgraph $\boldsymbol{G}_s$ of graph ***G*** will be used to extract the complete subgraph $\boldsymbol{G}_{cs}$, and then based on certain rules, the granular will be generated, which is a process of granular computing in this paper. Therefore, how to find all complete subgraphs in the graph $\boldsymbol{G}_s$ is very important, especially for the complex data sets. The algorithm for extracting all complete subgraphs is developed by the similarity matrix, and it can be described as follows:

// find all complete subgraphs

CG={ }; // ***CG*** denotes the set of complete subgraphs

for *i*=1 : *n* -2

if sum($\boldsymbol{M}(i,:)$)≠0

$\boldsymbol{X}_s$={ }; $\boldsymbol{X}_s$={x_j | j=i+1,…,n, e_{ij}≠0}; // $\boldsymbol{X}_s$ is a subset of $\boldsymbol{X}$

$\boldsymbol{X}_s$={x_i, $\boldsymbol{X}_s$};

$\boldsymbol{M}_s$ is a $|\boldsymbol{X}_s|\times|\boldsymbol{X}_s|$ similarity submatrix generated by $\boldsymbol{X}_s$;

tag=0; // tag is a counting sign;

in the $\boldsymbol{M}_s$, if ∀ e=0, tag=tag+1, and then store two samples relating to e in the position matrix $\boldsymbol{P}_{2\times tag}$; // $\boldsymbol{P}$ indicates the two samples have no connection, so the two samples can not exist in a same complete graph

if tag=0 // it means all objects connect with each other

$\boldsymbol{CG}$={{$\boldsymbol{X}_s$}, $\boldsymbol{CG}$};

else

$\boldsymbol{CG}_t$={}; $\boldsymbol{CG}_t$={$\boldsymbol{X}_s$}; // $\boldsymbol{CG}_t$ is used to store samples sets temporarily

for t=1 : tag

for k=1 : $|\boldsymbol{CG}_t|$

∃ $\boldsymbol{P}(t,:)\in\boldsymbol{CG}_t(k)$;

$\boldsymbol{CG}_t(k)$= $\boldsymbol{CG}_t(k)$- $\boldsymbol{P}(t,1)$;

and $\boldsymbol{CG}_t(|\boldsymbol{CG}_t|+1)$= $\boldsymbol{CG}_t(k)$+ $\boldsymbol{P}(t,1)$ - $\boldsymbol{P}(t,2)$;

end

end

$\boldsymbol{CG}$={$\boldsymbol{CG}_t$, $\boldsymbol{CG}$};

end

end

end

∀ $\boldsymbol{CG}(i)\in\boldsymbol{CG}(j)$, s.t. i, j=1,2,..., $|\boldsymbol{CG}|$, $j\neq i$, delete $\boldsymbol{CG}(i)$; // if containing relation of any two subsets in $\boldsymbol{CG}$ is existence, delete one subset

// the program for finding all complete subgraphs runs to completion

Based on the previous procedure, all complete subgraphs are stored in $\boldsymbol{CG}$. And for granular, it can be generated using the trivial and complete subgraphs. So for the trivial subgraphs, the granular is constructed as follows:

$\boldsymbol{Gr}$={}; // $\boldsymbol{Gr}$ is the set of granular

$\boldsymbol{Gr}$={{$\boldsymbol{x}_i$}| i=1,2,...,n, sum($\boldsymbol{M}(i,:)$)+sum($\boldsymbol{M}(:,i)$)=0}; // $\boldsymbol{x}_i$ constitutes a $\boldsymbol{G}_{ts}$ in the graph $\boldsymbol{G}$

For the complete subgraphs, the process of generating granular is relatively complicated. To solve the problem, three rules are proposed for coping with $\boldsymbol{CG}$. Firstly, the subset containing the most elements will form a granular, and the purpose is to avoid generating the granulars containing only one sample. Secondly, the subsets containing the same number of elements will depend on the weights, and the subset with higher weight value will form a granular. Furthermore, once a subset is selected as a granular, and some elements existing in other subsets will be deleted. Thirdly, if a subset only contains one element, it will also form a granular. The process will be repeated until each sample is assigned to a granular. The process of granular generation is graph partition. Therefore, the granular is constructed as follows:

while $|\boldsymbol{CG}|$>0

$\boldsymbol{CG}_t$={}; // $\boldsymbol{CG}_t$ is used to store complete subgraph with most elements temporarily

$\boldsymbol{CG}_t$={$\boldsymbol{CG}(i)$ | arg max($|\boldsymbol{CG}(i)|$), s.t. i=1,2,⋯, $|\boldsymbol{CG}|$};

if $|\boldsymbol{CG}_t|$=1

$\boldsymbol{Gr}$={$\boldsymbol{CG}_t$, $\boldsymbol{Gr}$}; // execute the first rule.

else

// $\boldsymbol{CG}_t$ is re-assign a complete subgraph with highest weight

$\boldsymbol{CG}_t$={$\boldsymbol{CG}_t$ (j) | arg max(sum($\boldsymbol{M}_{|\boldsymbol{CG}_t(j)|\times|\boldsymbol{CG}_t(j)|}$)), s.t. j=1,2,⋯, $|\boldsymbol{CG}_t|$}; // execute the second rule.

$\boldsymbol{Gr}$={$\boldsymbol{CG}_t$, $\boldsymbol{Gr}$};

end

the complete subgraph $\boldsymbol{CG}_t$ forming a granular will be deleted from $\boldsymbol{CG}$;

∀ $\boldsymbol{CG}(k)\cap\boldsymbol{CG}_t\neq\varnothing$, s.t. k=1,2,⋯, $|\boldsymbol{CG}|$, delete {$\boldsymbol{CG}(k)\cap\boldsymbol{CG}_t$} from $\boldsymbol{CG}(k)$;

∀ $\boldsymbol{CG}(l)\in\boldsymbol{CG}(k)$, s.t. l, k=1,2,⋯, $|\boldsymbol{CG}|$, $l\neq k$, delete $\boldsymbol{CG}(l)$; //eliminate containment relations

∀ $|\boldsymbol{CG}(k)|$=1, s.t. k=1,2,⋯, $|\boldsymbol{CG}|$, $\boldsymbol{Gr}$={$\boldsymbol{CG}(k)$, $\boldsymbol{Gr}$}, then delete $\boldsymbol{CG}(k)$;

end

So far, the whole process of generating granulars is ending, and it will be demonstrated by a representative case. There are 10 samples in this case, $\boldsymbol{X}$=[$\boldsymbol{x}_1$, $\boldsymbol{x}_2$, ..., $\boldsymbol{x}_{10}$]. The similarity matrix is given as follows:

$$\boldsymbol{M}_{10\times10}=\begin{Bmatrix}0&0&0&0&0&0&0&0&0&0\\0&0&0.97&0.95&0.97&0&0&0&0&0\\0&0&0&0.98&0.94&0&0&0.93&0&0\\0&0&0&0&0.96&0.93&0.97&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0.97&0&0.98&0.91\\0&0&0&0&0&0&0&0&0.98&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0.93\\0&0&0&0&0&0&0&0&0&0\end{Bmatrix}.$$

According to matrix operation, $\boldsymbol{x}_1$ constitutes a trivial subgraph, so it forms a granular automatically, i.e. $\boldsymbol{Gr}=\{\{\boldsymbol{x}_1\}\}$. The subgraph constituted by the other objects is used to extract all complete subgraphs. Take $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ as an example, for $\boldsymbol{x}_2$, $\boldsymbol{X}_s=\{\boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{x}_5\}$,

$$\text{so } \boldsymbol{M}_s=\begin{bmatrix}0&0.97&0.95&0.97\\0&0&0.98&0.94\\0&0&0&0.96\\0&0&0&0\end{bmatrix},$$

and all e in $\boldsymbol{M}_s$ are not equal to 0, that means tag=0, so $\boldsymbol{CG}=\{\{\boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{x}_5\}\}$. On the other hand, for $\boldsymbol{x}_3$, $\boldsymbol{X}_s=\{\boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{x}_5, \boldsymbol{x}_8\}$,

$$\text{so } \boldsymbol{M}_s=\begin{bmatrix}0&0.98&0.94&0.93\\0&0&0.96&0\\0&0&0&0\\0&0&0&0\end{bmatrix}.$$

In this example, on the contrary, tag=2, it can deduce that the position matrix $\boldsymbol{P}=\begin{bmatrix}\boldsymbol{x}_4&\boldsymbol{x}_8\\\boldsymbol{x}_5&\boldsymbol{x}_8\end{bmatrix}$. Accordingly, Figure 2 describes the process.

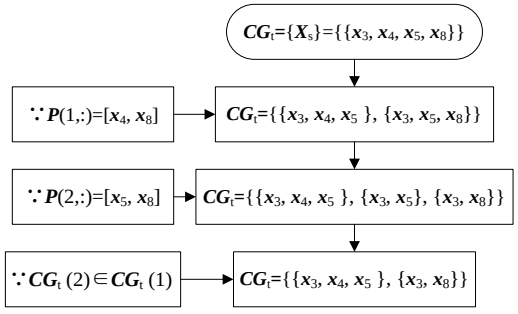


Figure 2. Demo diagram of extracting complete subgraphs

Hence, the $\boldsymbol{CG}=\{\{\boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{x}_5\}, \{\boldsymbol{x}_3, \boldsymbol{x}_8\}, \{\boldsymbol{x}_4, \boldsymbol{x}_6, \boldsymbol{x}_7\}, \{\boldsymbol{x}_6, \boldsymbol{x}_7, \boldsymbol{x}_9\}, \{\boldsymbol{x}_6, \boldsymbol{x}_9, \boldsymbol{x}_{10}\}\}$. Then Figure 3 can be used to describe the process of generating granulars. Finally, the 10 samples can generate 5 granulars, and $\boldsymbol{Gr}=\{\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{x}_5\}, \{\boldsymbol{x}_8\}, \{\boldsymbol{x}_6, \boldsymbol{x}_7, \boldsymbol{x}_9\}, \{\boldsymbol{x}_{10}\}\}$.

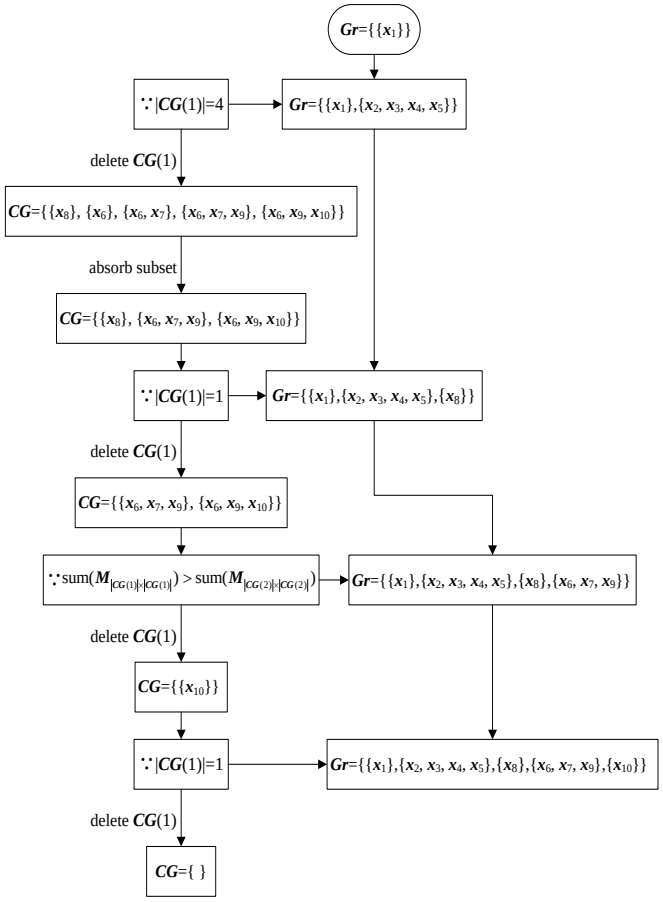


Figure 3. Demo diagram of generating granulars

It is shown in the case that a large number of similar samples form granulars can keep the diversity of granulars. Moreover, the granular formed by single sample maintains the original distribution of data. Of course, once more samples are processed, the intuitive graph partition depending on mankind is impossible to cope with the complicated big data. In this situation, huge calculation work can be only realized through matrix operation rules by the aid of advanced computer. So the granular computing using graph partition is achieved depending on matrix operation rules proposed above.

4. THE APPLICATION OF GRANULAR AND ITS EVALUATION

After generating the granular, it is unnecessary for all samples to participate in training, because the granular can replace the corresponding samples. It will reduce training samples and improve computational efficiency significantly. Therefore, application of the granular into fault diagnosis is important. Meanwhile, the evaluation system is designed for judging the spatial distribution under the changing of training data.

4.1. Application of Granular in Fault Diagnosis

In fault diagnosis, training samples satisfying a certain similarity can form a granular. Since many samples are close to each other in the fault feature space, its contributions to judging the fault pattern are basically similar. Accordingly, it is reasonable for a typical sample to represent all samples to participate in training; this will help to improve computational efficiency properly.

In this paper, the samples in same granular can be replaced by one data in the geometric center of the granular, and the data is designed as granular center (gc) sample. For the *i*-th granular $\boldsymbol{Gr}(i)$, the granular center sample can be calculated by:

$$\boldsymbol{x}_{\mathrm{gc}} = \frac{\sum_{j=1}^{|Gr(i)|} \boldsymbol{x}_j}{|Gr(i)|}, \tag{5}$$

where $\boldsymbol{x}_j$ is the sample in the *i*-th granular. Obviously, granular center sample is in the central position of a small neighboring area that formed by the samples in term of spatial distribution, and it can reflect the same fault pattern feature compared with other samples synthetically.

4.2. Evaluation for the Spatial Distribution Change

The spatial distributions of fault patterns are embodied in the original samples, and the samples in the same neighboring area may represent the same kind of fault pattern. Undoubtedly, any remarkably change to the original samples, *i.e.*, deleting, moving, replacing, can change the spatial distribution of fault pattern, which will influence the recognition of unknown samples. After replacing the samples, the change to original samples needs to be evaluated.

In this paper, the original samples center $\boldsymbol{x}_{\mathrm{oc}}$ is equal to the mean value of all samples, it can be calculated by the expression similar to Eq. (5). When the replacing is made, the statistics of the distances from $\boldsymbol{x}_{\mathrm{oc}}$ to both the original samples and the replacement samples must be changed, and the statistics should include Min, Mean, Max and STD (standard deviation) of distances. Furthermore, the statistics about original and new samples can respectively constitute two vectors $\boldsymbol{s}_{\mathrm{os}}$ and $\boldsymbol{s}_{\mathrm{ns}}$. Accordingly, based on the comparisons between $\boldsymbol{s}_{\mathrm{os}}$ and $\boldsymbol{s}_{\mathrm{ns}}$, their obvious difference means larger variation of spatial distribution; on the contrary, it means the smaller variation. That is the principle of a qualitative evaluation system.

In practice, the difference between $\boldsymbol{s}_{\mathrm{os}}$ and $\boldsymbol{s}_{\mathrm{ns}}$ can be measured quantitatively. In this study, a weighted relative error is designed as a criterion for evaluating the spatial distribution change. The weights of the statistics are distinguishing, and the changes of Min and Max are respectively single indicators. Relative small changes mean that the detailed distribution is maintained better. On the contrary, they don't necessarily imply the obvious change in distribution. Thus, the weights of the changes of Min and Max are smaller than the changes of Mean and STD, and the weight vector for the changes of Min, Mean, Max, and STD is $\boldsymbol{\omega} = [0.1, 0.4, 0.1, 0.4]$. Eventually, the weighted relative error (e_{WR}) can be calculated by:

$$e_{\mathrm{WR}} = \left(\sum_{i=1}^{|\boldsymbol{s}_{\mathrm{os}}|} \frac{|\boldsymbol{s}_{\mathrm{os}}(i) - \boldsymbol{s}_{\mathrm{ns}}(i)|}{\boldsymbol{s}_{\mathrm{os}}(i)} \cdot \boldsymbol{\omega}(i) \right) \cdot 100\%, \tag{6}$$

5. APPLICATION EXPERIMENTAL AND PERFORMANCE COMPARISONS

In order to verify the applicability of the proposed method, a typical 2-dimensional nonlinear data "twomoons" and a high-dimensional practical condition monitoring data of certain equipment, whose relative information are listed in Table 1, are used. The training and test numbers are selected randomly. Meanwhile, according to Eq. (1), similarity ranges of the two data are shown in Table 1.

5.1. Verify the Rationality of Parameter e_{WR}

Different similarity thresholds will generate coarser or finer granular. According to the previous analysis, granular computing for data compression must cause the spatial distribution change of data. Table 2 shows the situation of generating granular based on the different similarity thresholds. With the decreasing of similarity threshold, the number of granular is decreased. An extreme case is that all samples are classified as a granular which makes e_{WR}=100%. Especially, for 2-dimensional data, few granular do not necessarily imply the smaller e_{WR}. However, for the high-dimensional monitoring data, when the number of granular is reduced, the e_{WR} is also decreased. Generally, the changing trend of e_{WR} is increased with the decreasing of similarity threshold, as shown in Figure 4.

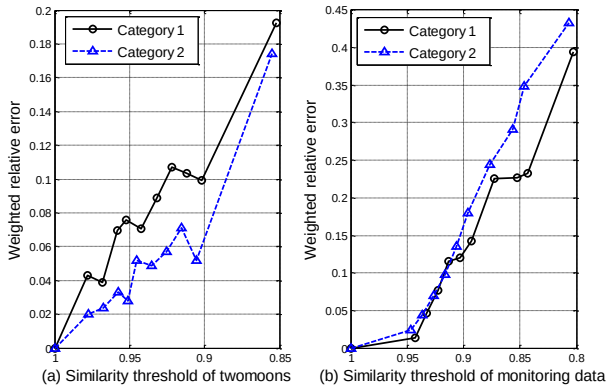


Figure 4. Change relationship between similarity threshold

and e_{WR}

Table 1. The relative information about data sets

Name	Dimensions	Categories	Train samples	Test samples	Similarity range
Twomoons	2	2	1002 (668+334)	500 (333+167)	[0.5679 0.9999] [0.5667 0.9993]
Monitoring data	10	2	1334 (667+667)	628 (333+295)	[0.3242 0.9754] [0.3406 0.9783]

Table 2. Situation of generating granular based on different similarity thresholds

Twomoons			Monitoring data		
Similarity threshold	Granular (sample) Number	e_{WR}	Similarity threshold	Granular (sample) number	e_{WR}
[0.9783 0.9777]	[154 106]	[4.30% 2.03%]	[0.9429 0.9464]	[595 598]	[1.39% 2.44%]
[0.9683 0.9677]	[87 74]	[3.90% 2.36%]	[0.9329 0.9364]	[522 554]	[4.61% 4.34%]
[0.9583 0.9577]	[64 51]	[6.94% 3.31%]	[0.9229 0.9264]	[438 496]	[7.70% 6.97%]
[0.9520 0.9510]	[47 39]	[7.57% 2.81%]	[0.9129 0.9164]	[366 442]	[11.60% 9.79%]
[0.9420 0.9451]	[38 34]	[7.07% 5.18%]	[0.9029 0.9064]	[307 377]	[11.96% 13.54%]
[0.9320 0.9351]	[32 25]	[8.86% 4.87%]	[0.8929 0.8964]	[246 319]	[14.26% 18.03%]
[0.9220 0.9251]	[23 23]	[10.71% 5.69%]	[0.8729 0.8764]	[162 237]	[22.59% 24.46%]
[0.9120 0.9151]	[21 18]	[10.32% 7.10%]	[0.8529 0.8564]	[104 182]	[22.64% 29.09%]
[0.9020 0.9051]	[17 15]	[9.92% 5.17%]	[0.8429 0.8464]	[91 156]	[23.25% 34.91%]
[0.8520 0.8551]	[11 11]	[19.23% 17.43%]	[0.8029 0.8064]	[42 95]	[39.44% 43.28%]

Furthermore, some spatial distributions about data set "Twomoons" will be drawn in a two-dimensional plane (Figure 5) for demonstrating the applicability directly. Figure 5 demonstrates that the detailed distributions of subgraphs are similar to the original spatial distribution. The larger weighted relative error is, the more obvious change in spatial distribution will be. Therefore, e_{WR} corroborates the conclusion reflected in Figure 5. Accordingly, the parameter e_{MR} is reasonable for measuring the change to the original samples.

5.2. Influence of GPDS on classification algorithms

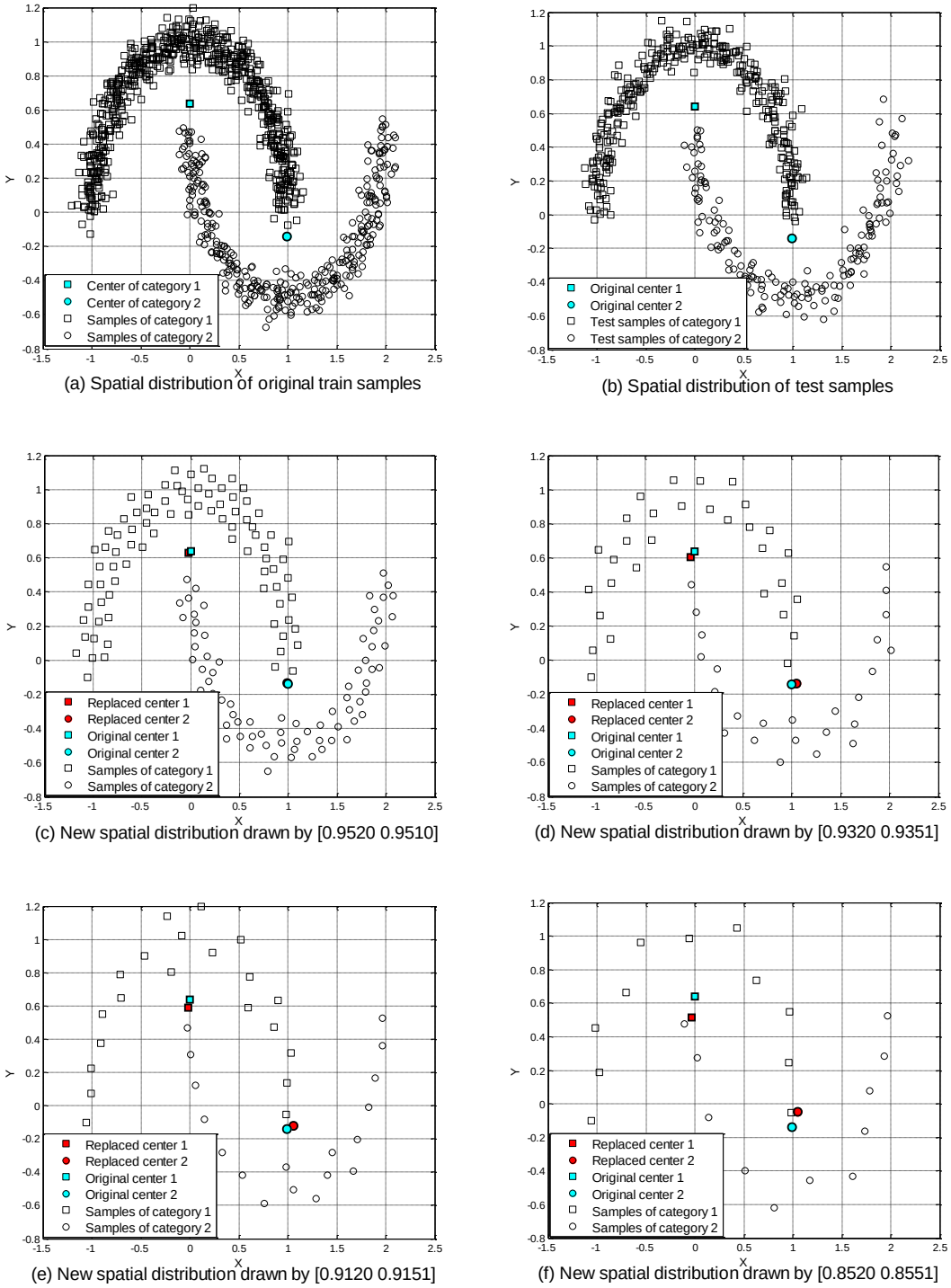
For keeping classification accuracy, the validity and rationality of granular computing for reducing computational cost will be verified. For classification algorithms based on iteration operations, the number of iteration is always influenced by the number of training samples. Support vector machine (SVM) optimized by particle swarm optimization (PSO) (Zheng, 2013), the learning vector quantization (LVQ) network (Biswal et al., 2014), the back propagation (BP) network (Ali et al., 2015), and the kernel multi-team competitive optimization (k-MTCO) (Zheng, 2016) are used to verify the applicability of granular computing. Some more detailed information can be found in the cited references. Especially, the population sizes of PSO and iteration number are set as 20 and 100. The population size of K-MTCO and the iteration number are set as 80 and 600, respectively; the error goals for BP and LVQ network are set as 0.001, and the iteration numbers for them are set as 1000 and 200, respectively. 2 data sets are scaled to be in interval [0, 1], which helps to reduce the search range and to improve the algorithm efficiency. Furthermore, all calculations are carried out under the same setting.

No free lunch (NFL) theorem indicates that any pattern classification algorithm cannot hold the superiority in its blood, it is impossible to be effective for all problems (D'Orsi et al., 2001). Because the classification principles and operation rules are completely different, the recognition accuracies calculated by the four algorithms are obviously differences. Figure 6 and Figure 7 demonstrate that the classification results and elapsed times of these four algorithms are based on two data sets. Meanwhile, Figure 7 show that the elapsed time is always decreased with the reduction of training samples.

For 2-dimensional nonlinear separable data, BP and SVM are superior to LVQ and K-MTCO. Under the precondition of rapid convergence, the computational cost of BP and LVQ are relatively lower. Moreover, the number of training samples has a significant influence on the accuracies of LVQ and K-MTCO, and the influence does not show any regularity with the change of training samples.

Essentially, Figure 6 indicates that the lower e_{WR} does not necessarily imply the higher accuracy. The difference of classification principles and the change of spatial distribution of data lead to the accuracy change with the quantity change of training samples. Especially, some deleted samples may reduce the misclassification of the algorithms. Therefore, how to determine the proper similarity thresholds is very important, and the future work will focus on it. However, the experimental results and

performance comparisons have shown the applicability of the proposed method. Generally, the granular computing based on GPDS can really improve the computational efficiency under the precondition of keeping classification accuracy, and it can be used to improve the efficiency of troubleshooting and shorten the maintenance period properly.



(a) Spatial distribution of original train samples

(b) Spatial distribution of test samples

(c) New spatial distribution drawn by [0.9520 0.9510]

(d) New spatial distribution drawn by [0.9320 0.9351]

(e) New spatial distribution drawn by [0.9120 0.9151]

(f) New spatial distribution drawn by [0.8520 0.8551]

Figure 5. Comparison of the change of spatial distribution

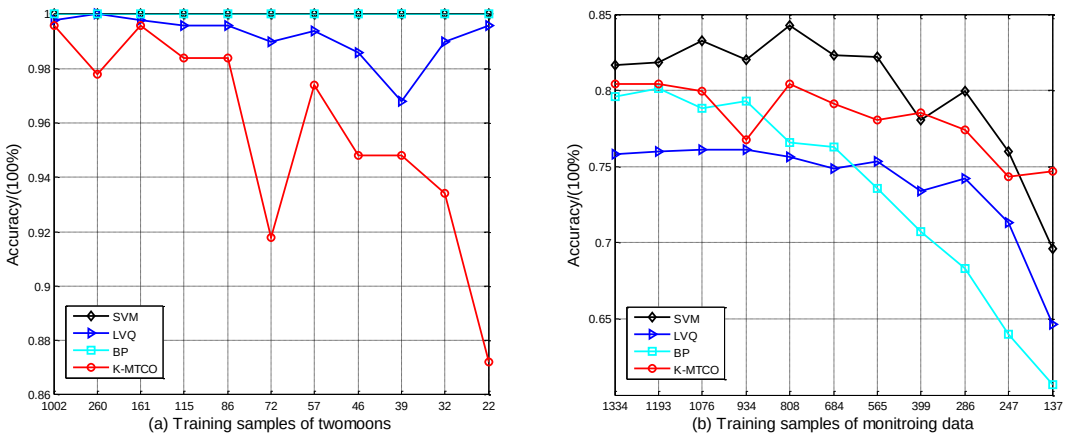


Figure 6. Relationship between the number of training samples and accuracy

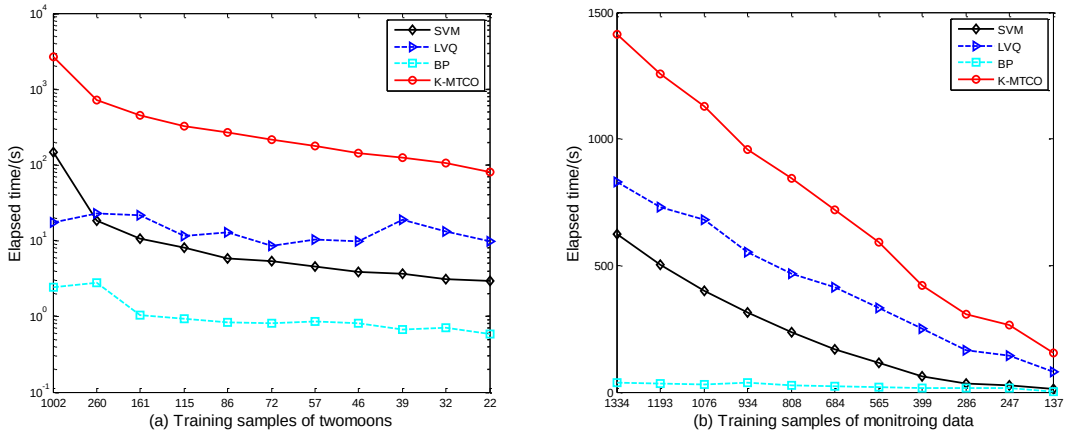


Figure 7. Relationship between the number of training samples and the elapsed time

6. CONCLUSIONS

In this paper, main conclusions obtained in the study are summarized as follows:

(1) The dimensionless similarity is proposed to overcome the influence of data with different dimensions, and similarity threshold determination method based on frequency distribution histogram is proposed to reduce the dependency on the experiences of expert. The results show that the proposed methods are rational and effective for engineering practice.

(2) The granular computing based on graph partition is designed to compress the original data; it can maintain the spatial detailed distribution of original data properly. Moreover, the parameter e_{MR} is proposed to measure the change of the original samples.

(3) The granular computing is used in the field of fault diagnosis. The computational efficiency is improved with the decreasing of training samples, and the proper compressed training samples can maintain the classification accuracy and a more reasonable result will be obtained. But, the influence of different number of training samples on the accuracies does not show any regularity. Therefore, how to determine the proper similarity thresholds is very important and thus the future work will focus on it.

NOMENCLATURE

$\boldsymbol{X}$	samples set
$\boldsymbol{x}$	a sample
d	dimension of a sample
r	range of a column
S	similarity

$\boldsymbol{GS}$	graph system
s	similarity threshold
$\boldsymbol{M}$	similarity matrix
$\boldsymbol{G}$	a graph
$\boldsymbol{G}_{\text{ts}}$	trivial subgraph
$\boldsymbol{G}_{\text{cs}}$	complete subgraph
$\boldsymbol{Gr}$	set of granular
$\boldsymbol{x}_{\text{oc}}$	original samples center
$\boldsymbol{s}_{\text{os}}$	the statistics about original samples
$\boldsymbol{s}_{\text{os}}$	the statistics about new samples
e_{WR}	weighted relative error

REFERENCES

Airteimoori, A., & Kordrostami, S. (2010). A euclidean distance-based measure of efficiency in data envelopment analysis. *Optimization*, vol.59, pp.985-996.

Ali, J. B., Fnaiech, N., Saidi, L., Chebel-Morello, B., and Fnaiech, F. (2015). Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration. *Applied Acoustics*, vol.89, pp.16-27.

Anjum, R. A., Zhang, Y. Q., Harrison, R. W. (2009). Granular decision tree and evolutionary neural SVM for protein secondary structure predictio. *International Journal of Computational Intelligence Systems*, vol2(4), pp.343-352.

Biswal, B., Biswal, M., Hasan, S., and Dash, P. K. (2014). Nonstationary power signal time series data classification using LVQ classifier. *Applied Soft Computing*, vol.18, pp.158-166.

D'Orsi, C. J. 2001. Computer-aided detection: there is no free lunch. *Radiology*, vol.221, pp.585-586.

Huang, M., Wang, J., & Liang, X. (2007). Genetic algorithm controlled by two thresholds for job scheduling problem. *Computer Integrated Manufacturing System*, vol.13, pp.329-332.

Li, Y. (2007). The computer networ routing research based on granularity computation intelligence. Hefei: *Anhui University*.

Liu, X. C. (1992). Entropy, distance measure and similarity measure of fuzzy sets and their relations. *Fuzzy Sets and Systems*, vol.52, pp.305-318.

Neto, A. M. (2014). Pearson's correlation coefficient: a more realistic threshold for applications on autonomous robotics. *Computer Technology and Application*, vol.5, pp.69-72.

Park, H. S., Pedrycz, W., Oh, S. K. (2009). Granular neural networks and their development through context-based clustering and adjustable dimensionality of receptive fields. *IEEE Transactions on Neural Networks*, vol.20, pp.1604-1016.

Pawlak, Z. (1991). Rough sets-theoretical aspects of reasoning about data. Dordrecht: Kluwer Academic Publishers.

Skowron, A., & Stepaniuk, J. (2001). Information granules: towards foundations of granular computin. *International Journal of Intelligent Systems*, vol.16, pp. 57-85.

Tang, Y. C., Jin, B., & Zhang, Y. Q. (2005). Granular support vector machines with association rules mining for protein homology prediction. *Artificial Intelligence in Medicion*, vol.35, pp.121-134.

von Luvburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, vol.17, pp.395-416.

Xiang, S. X., Nie, F. P., Zhang, C. S. (2008). Learning a mahalanobis distance metric for data clustering and classification. *Pattern Recognition*, vol.41, pp.3600-3612.

Ye, J. (2011). Cosine similarity measures for intuitionistic fuzzy sets and their applications. *Mathematical and Computer Modeling*, vol.53, pp.91-97.

Zadeh, L. A. (1996). Fuzzy logic-computing with words. *IEEE Transactions on Fuzzy System*, vol.4, pp.103-111.

Zadeh, L. A. (1997). Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets & Systems*, vol.90, pp.111-117.

Zhang, B., & Zhang, L. (1992). Theory and application of problem solving. North-Holland: Elsevisr Science Publisher.

Zhang, B., & Zhang. L. (2003). Theory of fuzzy quotient space (methods of fuzzy granular computing). *Journal of Software*, vol.14, pp.770-776.

Zhang, X. (2000). Research on granular support vector machine. Taiyuan: *Shanxi University*.

Zhang, Y. P., Zhang, L., & Wu, T. (2004). The representation of different granular worlds: a quotient space. *Chinese Journal of Computers*, vol.27, pp.328-333.

Zhao, X. L., & Yang, Y. (2007). Development status and application research on granular computing. *Computer Engineering*, vol.33, pp.101-103.

Zheng, B. (2013). Investigation on aeroengine maintenance level decision on PSO-SVM. *Journal of Propulsion Technology*, vol.34, pp.687-692

Zheng, B., & Gao, F. (2015). Fault diagnosis method based on S-PSO classification algorithm. *Acta Aeronautica et Astronautica Sinica*, vol.36, pp.3640-3651.

Zheng, B., Huang, H. Z., & Li, Y. F. (2015). Aeroengine fault diagnosis method based on optimized supervised kohonen network. *Journal of Donghua University*, vol.32, pp.1029-1033.

Zheng, B., Huang, H. Z., Xu, H. W., Meng, D. B., and Zhang, X. L. (2016). Multi-team competitive optimization algorithm and its application in bearing fault diagnosis. *2016 Annual Reliability and Maintainability Symposium*.

Zhong, W., He, J. Y., Harrison, R., Tai, P. C., and Pan, Y. 2007. Clustering support vector machines for protein local structure prediction. *Expert Systems with Application*, vol.32, pp.518-526.

BIOGRAPHIES

Bo zheng is currently a Ph.D. candidate in mechanical engineering at the University of Electronic Science and Technology of China. He received his B.S. degree in vehicle application engineering from Civil Aviation Flight University of China in 2010. His research interests include fault diagnosis and pattern recognition, and optimization.

Hong-Zhong Huang is a Professor of the School of Mechanical, Electronic, and Industrial Engineering, at the University of Electronic Science and Technology of China. He has held visiting appointments at several universities in the USA, Canada and Asia. He received a PhD degree in Reliability Engineering from Shanghai Jiaotong University, China and has published 200 journal papers and 5 books in fields of reliability engineering, optimization design, fuzzy sets theory, and product development.

Jie Zhou is currently a Ph.D. candidate in Mechanical Engineering at the University of Electronic Science and Technology of China. He received his B.S. degree in Mechanical Design, Manufacturing, and Automation from University of Electronic Science and Technology of China in 2013. His research interests include fatigue life prediction and fatigue reliability analysis.

Yanfeng Li is an associate Professor of the School of Mechanical, Electronic, and Industrial Engineering, at the University of Electronic Science and Technology of China. He received a PhD degree in Reliability Engineering from University of Electronic Science and Technology of China and has published 30 journal papers in fields of reliability engineering, system reliability modeling, dynamic fault analysis, and BN based modeling and probability inference.