Research on Diagnosis of AC Engine Wear Fault Based on Support Vector Machine and Information Fusion

Lei Zhang
Department of Electrical & Electronic Engineering, Henan University of Urban Construction, Ping Ding Shan , 467036, China
Email: hncjzhanglei@yeah.net

Yanfei Dong
Department of Electrical & Electronic Engineering, Henan University of Urban Construction, Ping Ding Shan , 467036, China
Email: dong yanzi@hncj.edu.cn

***Abstract*—Support Vector Machine (SVM) and information fusion technology based on D-S evidence theory are used to diagnose wear fault of AC engines. Firstly, based on a number of frequently used oil sample analysis methods for detecting engine wear fault, establish corresponding sub SVM classifier. The classifier can reflect the mapping relation between fault symptoms and fault types and achieve the result for a single diagnosis item. And then, use D-S evidence theory to make information fusion over result for a single diagnosis item so as to make fault diagnosis. With diagnosis of AC engine wear fault serving as example, example testing is performed. The result shows that in comparison with conventional methods, the combination of SVM and information fusion technology is fast and effective, suitable for diagnosis of AC engine wear fault.**

***Index Terms*—Support Vector Machine (SVM), AC engine, fault diagnosis, information fusion**

I. INTRODUCTION

As an effective tool for solving nonlinear issues, artificial neural network is extensively applied in the field of mechanical fault [1]. It is one more artificial intelligence (AI) technology applied in diagnosis of engine wear fault, following application of expert system. However, as neural network technology is merely a heuristic technology reliant on experience and lacks solid theoretical basis, and its learning process uses the empirical risk minimization principle, making it likely to tend to local minimum and weak in generalization. Moreover, the complexity of its algorithm is considerably subject to influence of the complexity of network structure and samples. These weak points have checked its further application and development in smart fault diagnosis. In recent years, Support Vector Machine (SVM) proposed by Vapnik has received extensive attention [2]. It is a new machine learning approach based on statistical learning theory and structure risk minimization theory and specific to small sample sizes. It features brief mathematical form, intuitive geometric interpretation and excellent learning performance and promotion ability. It can overcome the above-mentioned weakness of neural network and has become a new popular research field in the realm of machine learning. Additionally, it has found successful application in such realms as pattern recognition, regression analysis and function approximation [3] [4] [5].

On the other hand, in practical fault diagnosis, the method of oil sample analysis has become the major method for diagnosis of wear fault of AC engine. It includes ferrography analysis, spectral analysis, grain counting analysis method and physical and chemical analysis. However, one single method is invariably subject to certain limitation in terms of test accuracy, hence unsatisfactory accuracy. If diagnosis information of these analysis methods can be used to the greatest extent possible, with fusion diagnosis made to have results complement each other, the accuracy of fault diagnosis can be improved. SVM boasts rapid learning speed and strong generalization, while D-S evidence theory boasts a fairly good ability in processing of uncertain information [6]. Therefore, the present article proposes a fusion diagnosis technology for engine wear fault, where firstly, SVM classifier is used to realize diagnosis with each single method, and then based on D-S evidence theory, diagnosis results undergo fusion so as to improve accuracy and reliability of diagnosis.

II. SVM

SVM is a machine learning algorithm proposed by Vapnik in the 1990's. With its good theoretical background and structure risk minimization principle, SVM provides a brand-new direction for machine learning. At first SVM is used for solving pattern recognition issues. Subsets of training data chosen for the purpose of discovering decision rules with generalization ability are called support vectors. Best support vector

doi:10.4304/jcp.7.9.2292-2297

separation is equivalent to separation of all data. SVM involves from optimal hyperplane given linear separability, as shown in Figure 1. By way of fixing risks, minimizing confidence risk and mapping input space to high-dimension inner product space, SVM effectively avoids "curse of dimensionality". SVM obtains the global optimum by solving a quadratic programming problem of linear constraint, hence no local minima, and the Fast algorithm ensures the rate of convergence. Typical SVMs for classification are shown in Figure 2.

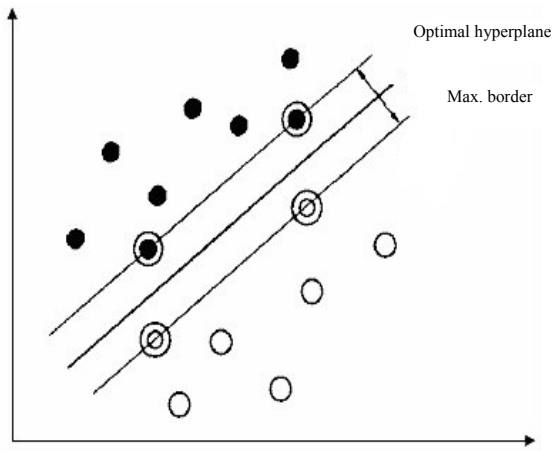

Figure 1 Optimal Hyperplane

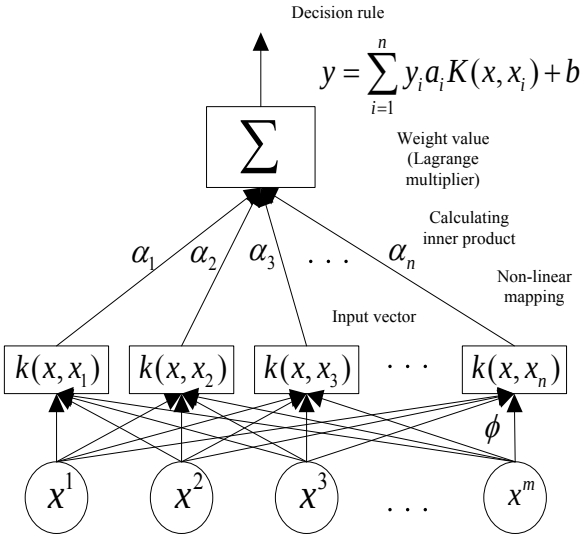

Figure 2 SVM Structure Network Diagram

A. Classification Algorithm of SVM

For given sample set $\{(x_i, y_i)\}_{i=1}^{N}$, $x_i \in R^m$ and $y_i \in \{\pm 1\}$, SVM firstly use nonlinear mapping $\phi: R^m \to R^n$ to map input vector to the high-dimension space. When data is separable in the high-dimension space, SVM constructs a maximum-interval separating classification hyperplane in the high-dimension space R^n: $(w \cdot \phi(x)) + b$, can be proven that w can be written as linear combination of ϕ: $w = \sum_{i=1}^{N} a_i y_i \phi(x_i)$; among that, a_i is the multiplier of Lagrange, and can be obtained by solving the following quadratic programming problem:

$$\underset{a}{Max} \quad \sum_{i=1}^{N} a_i - \frac{1}{2}\sum_{i,j=1}^{N} a_i a_j y_i y_j \left[\phi(x_i)\cdot\phi(x_j)\right]$$

$$s.t. \ \sum_{i=1}^{N} a_i y_i = 0 \qquad a_i \ge 0 \qquad (1)$$

It is known from theory of Reproduce Kernel Hilbert Space that: In the inner product of high-dimension space $\left[\phi(x_i)\cdot\phi(x_j)\right]$, a kernel function meeting Mercer conditions can always be found in the input space to enable $K(x_i, x_j) = \left[\phi(x_i)\cdot\phi(x_j)\right]$; therefore, there's no need to know the specific form of nonlinear mapping. Formula (1) can be adapted to the following form:

$$\underset{a}{Max} \quad \sum_{i=1}^{N} a_i - \frac{1}{2}\sum_{i,j=1}^{N} a_i a_j y_i y_j K(x_i, x_j)$$

$$s.t. \ \sum_{i=1}^{N} a_i y_i = 0 \qquad a_i \ge 0 \qquad (2)$$

Frequently used kernel functions include: Polynomial kernel $K(x,y) = [1+(x\cdot y)]^p$, $p = 1,\cdots,n$; Gaussian radial basis kernel $K(x,y) = \exp(-\|x-y\|^2 / 2\sigma^2)$ and two-layer feed-forward neural network kernel $K(x,y) = \tanh[k(x\cdot y)-\theta]$.

For points on the classification border, their corresponding $a_i > 0$ is called support vector points. The number of support vector points is generally smaller than sample size, and support vector points are closely related to generalization ability of classifiers. It is necessary to use Karush-Kuhn-Tucher condition to get the domain value b. Therefore, the classification decision function obtained is:

$$y = \mathrm{sgn}\left[\sum_{i=1}^{N} y_i a_i K(x, x_i) + b\right] \qquad (3)$$

When data can not be separated without error in the high-dimension space, SVM introduces non-negative slack variable $\xi_i \ge 0$ in constraints and solves the following quadratic programming to minimize wrong-classification error:

$$\min_{w,b,\xi} \ \frac{1}{2} w^T w + C(\sum_{i=1}^{N} \xi_i)$$

$$s.t. \ y_i(w^T x_i + b) \ge 1 - \xi_i \quad i = 1,\cdots,N \qquad (4)$$

Among that, slack variable $\xi = (\xi_1,\cdots,\xi_N)^T$ reflects the distance between the actual indication value y_i and SVM output. C is a marginal factor that reflects the balance between item 1 and item 2. Solving of Formula (4) can be converted to:

$$\underset{a}{Max} \quad \sum_{i=1}^{N} a_i - \frac{1}{2}\sum_{i,j=1}^{N} a_i a_j y_i y_j K(x_i, x_j)$$

$$s.t.\ \sum_{i=1}^{N} a_i y_i = 0 \quad 0 \le a_i \le C \qquad (5)$$

To solve Formula (5), likewise KKT condition can be used to obtain domain value and get the decision function (3).

B. Algorithm of Multi-class SVM Classification

Basic principle of multi-class SVM is converting multi-class issues into combination of two-class issues. In the present article, pair classification is adopted, i.e. specific to N-term classification issue, build $N(N-1)/2$ SVM sub-classifier(s); train SVM to separate between two classes, using the following algorithm:

Supposed training data sample belongs to Class k and Class l; the 2-term classifier can be converted into a multi-term classifier:

$$\min_{w^{kl},b^{kl},\xi^{kl}} \frac{1}{2}(w^{kl})^T w^{kl} + C(\sum_{i=1}^{N} \xi_i^{kl})$$

$$s.t. \begin{cases} [(w^{kl})^T \phi(x_i)] + b^{kl} \ge 1 - \xi_i^{kl} & x_i \in k \\ [(w^{kl})^T \phi(x_i)] + b^{kl} \le -1 + \xi_i^{kl} & x_i \in l \end{cases} \quad \xi_i^{kl} \ge 0 \qquad (6)$$

After using multi-term classifier to train data samples, when new samples are tested, input new samples to each trained SVM classifier to make classification recognition, thus obtaining a class each time. And then, make statistics on the result of $N(N-1)/2$ classifications; the class covering the most class number will become the class for the new sample. Multi-term classifier constructed with such an approach features less single SVM training size, easy realization using programming, accurate result and balanced training data.

III. Fusion Diagnosis Based on D-S Theory

A. Basis of D-S Evidence Theory

D-S theory uses "recognition frame work Θ" to describe the aggregate of all elements constituting a whole assumed space; it is made up of mutually exclusive and exhaustive elements. It defines a set function m: $2^{\Theta} \to [0,1]$, meeting: (1) $m(\varphi)=0$; (2) $\sum_{A \subset \Theta} m(A) = 1$; m is called basic probability assignment function (BPA) on the recognition framework Θ; m(A) shows the extent of accurate confidence of evidence against A. For any proposition set, D-S theory also proposes concept of confidence function, i.e.:

$$\text{Bel(A)} = \sum_{B \subset A} m(B), \quad (\forall A \subset \Theta) \qquad (7)$$

That is to say, confidence function of A is the sum of confidence values of each sub-set. Bel function is also called lower limit function, and reflects full confidence in A. However, for confidence in one specific proposition, description with confidence function alone is not all-around enough because Bel(A) can not reflect the extent of suspicion on A. Therefore, a likelihood function is introduced, defining that: Pl(A)＝1—Bel(-A), $(\forall A \subset \Theta)$; Pl function is also called upper limit function or irrefutable function, reflecting confidence in that Proposition A is not false. To make it easy for understanding, it is an uncertainty metrics reflecting whether Proposition A probably will be true. Obviously: Pl (A)$\ge$ Bel(A), to all A $\subset \Theta$.

B. D-S Combination Rule

For the same recognition framework, depending on varied evidence, different BPA functions which are independent of each other will be obtained. Supposed m_1 and m_2 are two basic BPA functions on the same recognition framework, and m_1 and m_2 can be combined into a new BPA function $m_1 \oplus m_2$. The corresponding confidence function is expressed as $Bel_1 \oplus Bel_2$; and according to definition of confidence function, $Bel_1 \oplus Bel_2$ calculation can be made by using $m_1 \oplus m_2$.

Define $m(A) = \frac{1}{N} \sum_{B \cap C \ne \phi} m_1(B) \cdot m_2(C)$ as the BPA function of $m_1 \oplus m_2$.

In this formula, $m(A)$ is the *mass* function on Θ; $N = 1 - \sum_{B \cap C \ne \phi} m_1(B) \cdot m_2(C) > 0$. When $N = 0$, $m_1 \oplus m_2(A)$ makes no sense, representing that the two BPA functions of $m_1(B)$ and $m_2(C)$ are in full conflict and can not be combined.

IV. The Combination of SVM and D-S Theory Evidence

D-S evidence theory is an important method in the field of multi-sensor information fusion, but its advantage is not fully utilized because is BPA is difficult to obtain. SVM is a new learning algorithm based on the statistical learning theory. However, its hard decision output does not adequately facilitate multi-sensor information fusion. In order to apply SVM to information fusion, a two-class SVM with BPA output is proposed. By analyzing the essence and deficiency of the Platt's model , the BPA is obtained through use of the lower bound of the SVM precision to weight the Platt' s probability model , which achieves the combination of SVM and the evidence theory in the information fusion.

The standard SVM output is the {1, 1}, belonging to a judge output, not probability output, so that cannot be as the evidence of BPA theory. SVM model of fault diagnosis is more than a class classification problems, "one-versus-one" many kinds of support vector machine (SVM) is an effective way to solve this problem, one is to use the ballot to judge fault mode.

For the first L sensors, in recognition framework Θ = {F,},s = 1 , 2 , …, t, each "one-versus-one" multi-class support vector machine (SVM) output corresponding to a evidence, if the total number of votes each V (F,) and total votes t (t-1) / 2 than can be obtained with probability

output, based on this, the structure, the basic allocation function:

$$m_1(F_S) = \frac{2V(F_S)}{t(t-1)}, A \neq \Phi, s = 1,2,\cdots,t \qquad (8)$$

If using a single source of data to fault diagnosis, the probability of the largest such as sample points x belongs of fault; If use evidence theory information fusion of data to fault diagnosis, the "one-versus-one" multi-class SVM output is a evidence.

V. Model of Fault Diagnosis System

Respect to four basic oil sample analysis technologies for AC engine wear faults, namely ferrography analysis, spectral analysis, grain counting analysis and physical and chemical analysis, the present article firstly takes some frequently occurring fault sets as a fault domain [7][8][9], and then specific to each oil sample analysis method, establish sub support vector classifiers corresponding to the fault sets, thus realizing mapping of each fault symptom and fault type and subsequently finishing preliminary diagnosis of fault types. After undergoing information fusion, preliminary information is sent to a classifier SVC for fault classification, i.e. realizing final classification of decision diagnosis.

As various oil sample analysis methods produce varied diagnosis data and dimensions, subsequent fusion Ⅶwill encounter some inconveniences. Therefore, pretreatment on original symptom data will be made to convert them into Boolean values 0 and 1. The basis for the conversion is: compare original data obtained by way of various methods against their corresponding standard limit values; if they are within the normal range, 0 is obtained; if not, 1 is obtained.

The realization of decision information fusion is a process within the same recognition framework, where different evidence bodies are combined into a new evidence body. The specific fusion method is as follows: supposed $S = \{s \mid s = 1,2,\cdots,q\}$, $\Theta = \{j = 1,2,\cdots,p\}$; for symptom domain s, local information fusion leads to the result of Model j: $ms(j)$; supposed the confidence factor of symptom domain s being used for local diagnosis is $R(s)$, $R(s) \in (0,1)$, the BPA function is defined as follows: $mass(j) = ms(j) \times R(s)$, where $j = 1,2,\cdots,p$. $mass(\theta) = 1 - R(s)$.With that determined, it is possible to use combination rules of D-S evidence theory to make global information fusion, judge probability of occurrence of various faults, and eventually obtain the final diagnosis of various fault patterns.

In order to directly show the fault type, the final decision diagnosis uses fault classifiers. As SVM is a 2-value classifier, fault classification of multiple fault types requires construction of multi-term classifier to enable fault diagnosis classification.

The model of AC engine wear fault diagnosis system based on support vector and D-S evidence theory fusion is shown in Figure 3.

The diagnosis process using the model of AC engine wear fault diagnosis system based on support vector and D-S evidence theory fusion is as follows:

(1) Preparation of training set and testing set. Use known diagnosis results as sample data for respectively training their own classifiers.

(2) After fusion of output of various classifiers by way of D-S evidence theory, take the fusion result as the input for fault classifier to train decision rules of SVM.

(3) Use testing set to test diagnosis system.

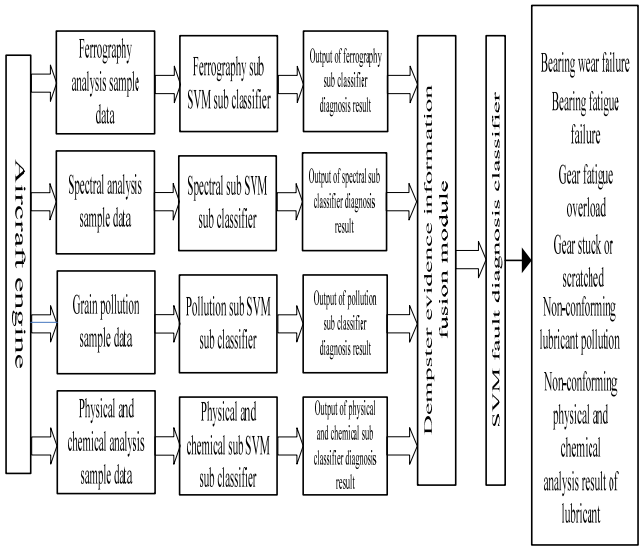

Figure 3 Model of AC Engine Wear Fault Diagnosis System

VI. Analysis of Fusion Diagnosis Instance

A. Diagnosis Instance

To verify the validity of the diagnosis method based on SVM and evidence theory fusion, the present article takes 6 frequently occurring wear faults of AC engine as the fault domains of the diagnosis. The 6 wear faults are: bearing wear failure (F_1), bearing fatigue failure (F_2), gear fatigue overload (F_3), gear stuck or scratched (F_4), non-conforming lubricant pollution (F_5) and non-conforming physical and chemical analysis result of lubricant (F_6). In this way, the recognition framework of the evidence theory $\Theta = \{F_1, F_2, F_3, F_4, F_5, F_6\}$ is built. Therefore, number of output type of the four sub classifiers built specific to the fault domains - ferrography analysis, spectral classifier, grain counting classifier method and physical and chemical classifier is 6. Regarding ferrography analysis, its input vector is the percentage of abrasive grain of elements of various types, ①globular abrasive grains in large number (S_{F1}); ②layered abrasive grain in large number (S_{F2});③fatigued abrasive grain in large number (SF3);④cutting abrasive grain in large number (SF4);⑤severely slippery abrasive grain in large number (S_{F5});⑥ red oxide abrasive grain in large number (S_{F6}); ⑦ black oxide abrasive grain in large number (S_{F7}); therefore, the input node number of ferrography sub classifier is 7. For spectral sub classifier,

choose concentrations of elements of Fe, Cr, Ni, Mo, Cu, V, Zn, Al and Ti as the original data for spectral analysis (for other machinery, depending on structure and material of abrasion parts, elements chosen may vary). After pretreatment, the spectral data are turned into: ① non-conforming Fe element concentration (S_{S1}); ②non-conforming Cr element concentration (S_{S2}); ③non-conforming Ni element concentration (S_{S3}); ④non-conforming M_o element concentration (S_{S4}); ⑤non-conforming V element concentration (S_{S5}); ⑥non-conforming C_u element concentration (S_{S6}); ⑦non-conforming Zn element concentration (S_{S7}); ⑧non-conforming Al element concentration (S_{S8}); ⑨Non-conforming Ti element concentration (S_{S9}), so input node number of spectral sub classifier is 9; for grain counting sub classifier, as number of grain with specific size level can not correspond to fault pattern of engine, it is only possible to arrive at the conclusion whether the oil sample pollution is non-conforming, i.e. the classifier input vector is: ①non-conforming pollution (S_{C1}), so for grain counting sub classifier, the input node number is 1; for physical and chemical analysis sub classifier, the input vector is: ①non-conforming kinematic viscosity (S_{P1}); ②non-conforming impurity content (S_{P2}); ③non-conforming other physical and chemical specifications(S_{P3}). So the input node number can be set as 3. Afterwards, train all the sub classifiers and fault diagnosis sub classifiers of the system.

To verify the validity of the algorithm herein, an example is provided for verification. Supposed fault symptom data of ferrography analysis are {0,0,0,1,0,0,0}; supposed fault symptom data of spectral analysis are {0,1,0,0,0,0,0,0,0}; fault symptom data of pollution analysis are {1}; fault symptom data of physical and chemical analysis are {0,0,1}; Single-item diagnosis results of each sub classifiers are shown in Table 1.

TABLE I.
SINGLE-ITEM DIAGNOSIS AND FUSION DIAGNOSIS RESULTS

	Ferrography analysis	Spectral analysis	Pollution analysis	Conflict degree	Fusion result
Bearing wear failure	0.8	0.9	1	0	0.972
Bearing fatigue failure	0.0001	0	0	0	0
Gear fatigue overload	0	0	0	0	0
Gear stuck or scratched	0.8	0.1	1	0	0.6626
Non-conforming lubricant pollution	0.6	0	1	0	1
Non-conforming physical and chemical analysis result of lubricant	0.0001	0	1	1	1

TABLE II
INSTANCE OF SMALL CONFLICT FUSION

	Ferrography analysis	Spectral analysis	Pollution analysis	Conflict degree	Fusion result
Bearing wear failure	0.8	0.9	1	0.28	0.972
Non-bearing-wear failure	0.2	0.1	0	0.28	0.028

TABLE III
Instance of Large Conflict Fusion

	Ferrography analysis	Spectral analysis	Pollution analysis	Conflict degree	Fusion result
Gear stuck or scratched	0.8	0.1	1	0.92	0.6626
Non-gear-stuck-or-scratched	0.2	0.9	0	0.92	0.3373

B. *Analysis*

(1) Influence of Kernel Function and Its Parameters on Diagnosis

The article uses 2 types of kernel functions (i.e. polynomial kernel and Gaussian radial basis kernel) to compare influence of varied kernel functions on sub SVM classifier and fault diagnosis classifier. K-based cross validation and grid search method are used to choose from 7 parameters of SVM, where: $\varepsilon = 0.02$. The result shows that, δ^2 of Gaussian radial basis kernel produces better result in the range of $[1,5]$, but if $d \le 3$, the classification abilities of polynomial kernel and Gaussian radial basis kernel are close to each other, though the calculation speed of the later is a bit faster (0.06). Therefore, radius basis kernel is used herein.

(2) Research on Comparison between SVM and Neural Network Models

To make comparison between the diagnosis ability of neural network and SVM, here BP neural network is used to make diagnosis on test data. Structure of BP network is 7-12-6, 9-12-6, 1-12-6 and 3-12-6; the allowance is 0.001. The result shows that though BP network can also find out the wear fault, the confidence degree is only 0.57, lower than the confidence degree of SVM diagnosis. This also shows generalization ability of BP network is not as strong as SVM.

VII. Conclusion

(1) The present article applies SVM to research of diagnosis of AC engine wear faults, and the result shows that: SVM features strong generalization ability in issues with small sample size, particularly applicable to diagnosis of AC engine faults where obtaining of wear fault samples is difficult.

(2) Firstly, use sub SVM to make preliminary diagnosis; then use D-S evidence theory to make decision fusion on the diagnosis result. In this way, relatively weak diagnosis decision will support relatively strong diagnosis decision in a more effective way, and such issues as diversification and multiple classes of information in practical projects can be solved.

(3) The combination of SVM and information fusion technology based on D-S evidence theory can realize effective diagnosis of wear faults of AC engines.

References

[1] HUANG Xianghua, DING Yi. Engine sensor fault diagnosis based on geometric pattern recognition[J]. Acta Aeronautica et Astronautica Sinica, 2006, 27 (6) : 1018-1022 .

[2] Feng Z G, Shida K, Wang Q. Sensor fault detection and data recovery based on LS-SVM predictor [J] . Chinese Journal of Scientific Instrument, 2007, 28(2) : 193-197 .

[3] CHEN Tian, SUN Jianguo, HAO Ying. Neural network and dempster shafter theory based fault diagnosis for aero engine gas path [J]. Acta Aeronautica et Astronautica Sinica, 2006, 27(6) : 1014-1017 .

[4] Kang Jian li, Lu Yaping , Zhou Yinsheng . Wear Particle Recognition with Improved BP Algorithm [J]. Lubrication Engineering , 2005 (2) : 149-150.

[5] CHEN Guo. Intelligent fusion diagnosis of aeroengine wear faults [J] . China Mechanical Engineering, 2005, 16 (4) : 299-302.

[6] ZHU Daqi, YU Shenglin. Data fusion algorithm based on D-S evidential theory and its application for circuit fault diagnosis [J]. Acta Electronica Sinica, 2002, 30 (2) : 153-155.

[7] HUANG Weibin, HUANG Jinquan. On board self-tuning model for aero-engine fault diagnostics [J] . Journal of Aerospace Power, 2008, 23(3) : 580-584

[8] Wang Weihua , Yin Yonghui, Wang Chengtao . Wear Debris Recognition System Based on Radius Basis Function Network[J] . Tribology , 2003 , 23 (4) : 340 - 343 .

[9] LI Yanjun , ZUO Hongfu , WU Zhenfeng , etal. Wear Particles Identification Based on Dempster-shafer Evidential Reasoning [J]. Journal of Aerospace Power , 2003 , 18 (1) : 114- 118 .

Lei Zhang He Nan Province, China. Birthdate: November, 1979. is Electric Information Engineering Ph.D., graduated from School of Electric Information Engineering, Jiangsu University. And research interests on data mining, complex networks.

He is a senior lecturer of Dept. Department of Electrical & Electronic Engineering, Henan University of Urban Construction.

Yanfei Dong He Nan Province, China. Birthdate: Feb, 1976. is graduated from School of Electric Information Engineering, Wuhan University. And research interests on data mining, complex networks,.

She is a associate professor of Dept. Department of Electrical & Electronic Engineering, Henan University of Urban Construction.