Fault Diagnosis Method Based on a New Supervised Locally Linear Embedding Algorithm for Rolling Bearing

HONGFANG YUAN [1], XUE ZHANG [2], YANGYANG[3], HUAQING WANG [3*]
[1] College of Information Science and Technology, Beijing University of Chemical Technology, Chao Yang District, Beijing, 100029, P.R.CHINA
[2] School of Mechanical & Electrical Engineering, Beijing University of Chemical Technology, Chao Yang District, Beijing, 100029, P.R.CHINA
wanghq_buct@hotmail.com

Abstract: In view of the complexity and nonlinearity of rolling bearings, this paper presents a new supervised locally linear embedding method (R-NSLLE) for feature extraction. In general, traditional LLE can capture the local structure of a rolling bearing. However it may lead to limited effectiveness if data is sparse or non-uniformly distributed. Moreover, like other manifold learning algorithms, the results of LLE and SLLE depend on the choice of the nearest neighbors. In order to weaken the influence of the random selection of the nearest neighbors, R-NSLLE, a supervised learning method, is used to find the best neighborhood parameter by analyzing residual. In addition, a nonlinear measurement based on SLLE is proposed as new criterion. In this paper, the original feature set is obtained through singular value decomposition in the phase space reconstructed by the C-C method. R-NSLLE is used for nonlinear dimensionality reduction, which can further extract fault features. Following this, R-NSLLE is compared with other nonlinear methods of dimensionality reduction, such as SLLE, LLE, LTSA and KPCA. The effectiveness and robustness of R-NSLLE have been verified in the experiment, and the accuracy and silhouette coefficient of the proposed method have been further discussed. These show that this feature extraction method, which is based on R-NSLLE, is more effective and can identify the intrinsic structural of rolling bearing even when there is a little fault.

Key-Words: Phase Space Reconstruction, Manifold Learning, SLLE, LLE, Rolling Bearing, Fault Diagnosis, Feature Extraction

1 Introduction

During rolling bearing fault diagnosis, because vibration signals contain a wealth of information, bearing states can be effectively identified and classified by analysing and processing the vibration signal [1]. Traditional fault diagnosis methods often adopt the time-frequency signal as the general fault characteristics, and the short-time Fourier transform, and wavelet transform are more commonly used time-frequency analysis methods. The multisensor data fusion method collects datasets from different channels in order to achieve signal processing, which helps improve the accuracy of fault diagnosis. For complex, nonlinear, and non-Gaussian signals, higher order spectrums can be used to extract fault features. Before, rolling bearing fault diagnosis was built on the basis of one-dimensional signals. However, one-dimensional nonlinear bearing signals collected by the sensor are essentially a one-dimensional projection of the entire mechanical system during high-dimensional operations, which does not fully characterise complete fault information. According to the theory of dynamics, some features in the phase space remain unchanged [2]. So, we attempted to extract the bearing features from those of the phase space in order to find more fault information.

As a good method for nonlinear dimensionality reduction, manifold learning has paid great attention to many researchers. It was put forward in a famous magazine "Science" in 2000, and the basic idea of it is that if the data sampled from high-dimensional data manifolds, there must be a mapped low-dimensional smooth manifold in the European space [3]. Manifold learning has obtained very good application in face recognition and data mining. As more scholars on the research of the manifold learning algorithm, manifold learning has already applied in mechanical fault diagnosis, such as signal de-noising, feature extraction and data dimension reduction, etc. Wang Guanwei systematically studied the application of manifold learning in mechanical fault diagnosis [4]; Wang lei’s doctoral dissertation proposed a signal de-noising method using maximum variance on manifold learning algorithm (MVU); some people combined manifold learning with SVM algorithm for fault diagnosis of mechanical and electronic systems [5]; Literature [6] combined mode decomposition (EMD) with supervised the locally linear embedding (SLLE) put forward a new method of data mining; LTSA was used in diesel engine wear fault feature extraction obtained a better result in Literature [7]. Based on the understanding of the above methods, this paper compares a variety of different algorithms of manifold learning. Though each method has its own advantage, this paper chooses LLE as a suitable method for feature extraction of rolling bearing. LLE [8] is a kind of manifold learning, which is a new unsupervised learning method; it is to find the nature low-dimensional structure form original high-dimensional data in the observation space, using the weights of neighbour points to construct the output vectors [9-11]. In recent years, a lot of researchers are working on the improved method of this manifold learning algorithm. Wang combined the locally linear embedding (LLE) with the kernel Fisher for rolling bearing fault analysis [12]. Li Benwei proposes a new machinery fault diagnosis approach based on supervised locally linear embedding projection (SLLEP) [13]. However, many researchers focus on the combination of LLE with other algorithms, but do not study pure algorithms. This is of limited effectiveness when data is sparse or non-uniformly distributed. Besides, the results of LLE and SLLE depend on the choice of the nearest neighbours. Therefore, in order to increase the practicality and robustness of the algorithm, we proposed an improved supervised locally linear embedding (R-NSLLE), which is a nonlinear supervised learning method that can find the best neighbourhood parameters by analysing residual.

This paper presents an improved C-C method that integrates supervised locally linear embedding (R-NSLLE) as a new extraction method for bearing fault signal characteristics. By reconstructing the phase space, the original one-dimensional signal is mapped into a high-dimensional phase space. From this we can obtain the original feature set via singular value decomposition [14]. Next, R-NSLLE is used for nonlinear dimensionality reduction, and compared with another nonlinear dimensionality reduction method called KPCA [15, 16], as well as other manifold methods such as SLLE, LLE and LTSA [17]. The experimental results show that this feature extraction method based on R-NSLLE works better and can foind the intrinsic structure of rolling bearings.

The main flowchart is shown in Fig.1:

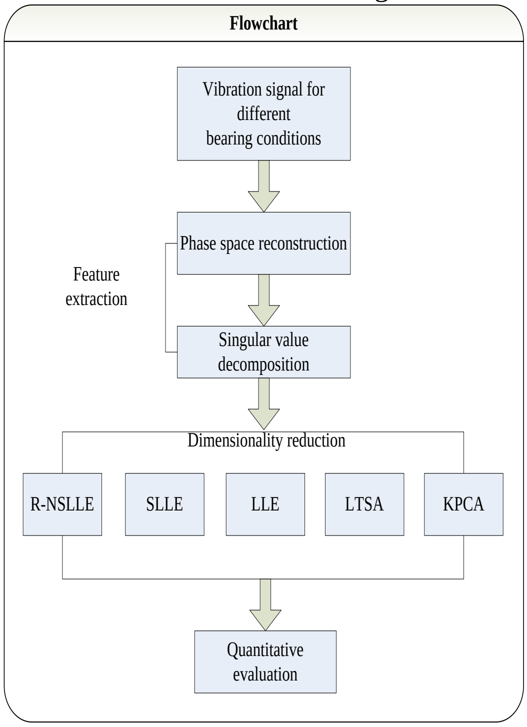


Fig.1 Flowchart of proposed rolling bearing fault diagnosis method

The rest of this paper is arranged as follows. Section 2 describes the feature extraction method in detail. Section 3 introduces LLE and SLLE, and then put forward a new algorithm called R-NSLLE. Furthermore, R-NSLLE is compared with SLLE, LLE, LTSA and KPCA by practical applications. Section 4 presents the quantitative evaluation of algorithm. Finally, conclusions are drawn in Section 5.

2 Feature extraction

2.1 Experimental data

The test data set used in this work was acquired from Diagnosis and Self-healing Engineering Research Center of Beijing University of Chemical Technology. The experimental platform is shown in Fig.2. The analyzed data consist of four different types based on the rolling element bearing are shown in Fig.3, including healthy, inner-race defect, outer-race defect, and rolling-element defect [18]. The bearing type is NTN204. And we set the sampling frequency 100 kHz with rotating speed 900r/min.

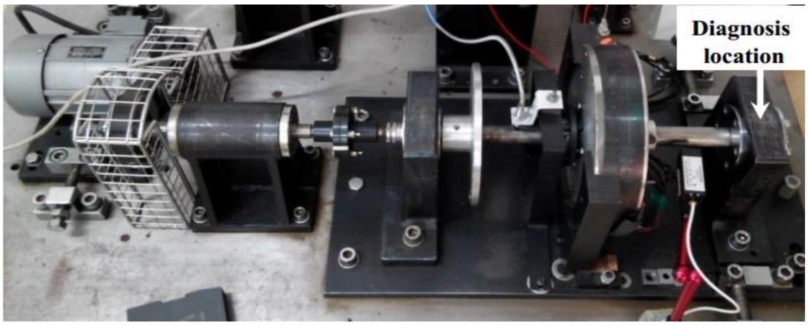


Fig.2 Experimental platform

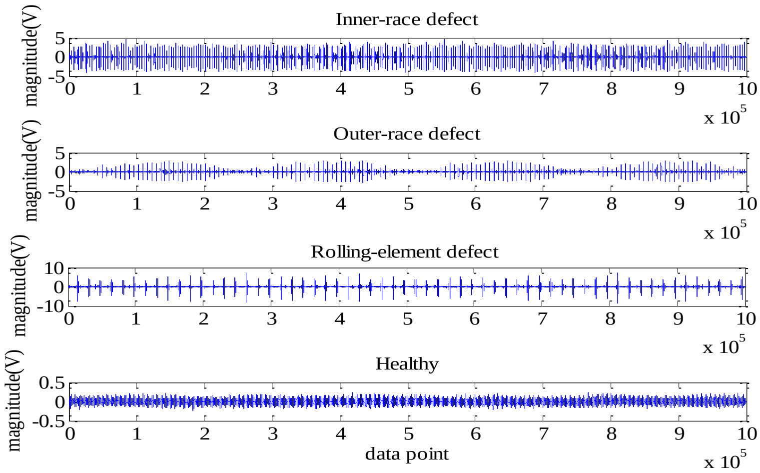
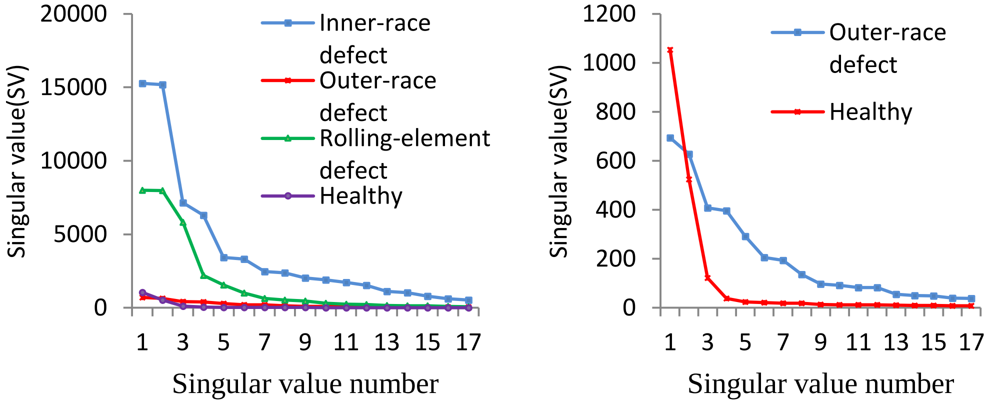


Fig.3 Bearing vibration signal

Table 1

The best time delay and embedding dimension of improved C-C method

Sample defect type	Time delay	Embedding dimension
Inner-race defect	3	11
Outer-race defect	3	19
Rolling-element defect	3	17
Healthy	3	17



(a) (b)

Fig.4 The singular value feature set

According to Fig.3, bearing characteristics under different fault status is not obvious, the difference, especially at where the impulses occurred, cannot be identified from the waveforms. Therefore, feature extraction becomes extremely important.

2.2 Improved C-C method

C-C method was put forward by H.S.Kim [19] and others in 1999; correlation integral was used to estimate the time delay and embedding window. However, there are some deficiencies of C-C method. The correlation integral defined generally uses the infinite norm, only considering the influence of the largest one-dimensional vector [20, 21]. There are some differences in calculating correlation integral for the improved algorithm. In this paper, two-norm is used instead of the infinite norm. In this way, the improved algorithm can reflect the effect of each dimension, which can well reflect the relevant characteristics of the original sequence. For different fault signals, we find the best time delay and embedding dimension by C-C method. The results are placed in Table 1 for comparison.

By reconstructing the phase space we can get a high-dimensional data feature set, it carries a lot of bearing fault information. Besides, for the convenience of the original feature set compression, take a unified embedding dimension m a value of 17, the time delay τ a value of 3 under four different states of the bearing vibration signal. Here the original signal x_t with N data points, the ith phase point vector in the m-dimensional phase as below:

$$X_i^m = [X_i,\ X_{i+\tau},\ \cdots,X_{i+(m-1)\tau}] \quad (1)$$

The constructed phase point vectors of m × n is given as:

$$\begin{bmatrix} X_1^m \\ X_2^m \\ \vdots \\ X_i^m \\ \vdots \\ X_n^m \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \\ x_2 & x_3 & \cdots & x_{m+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_i & x_{i+1} & \cdots & x_{i+(m-1)} \\ \vdots & \vdots & \vdots & \vdots \\ x_n & x_{n+1} & \cdots & x_n \end{bmatrix}^T = \begin{bmatrix} PS^1 \\ PS^2 \\ \vdots \\ PS^i \\ \vdots \\ PS^m \end{bmatrix} \quad (2)$$

Where $PS^j(j = 1,2,\cdots,m)$ denotes a vector with the meaning of time series [22].

2.3 Singular value decomposition

Take each 50 groups of data from four defect states on the rolling element bearing, thus the original data set reconstructed by improved C-C is 200*17 just as it shown in equation (2). After singular value decomposition [23, 24] we can get the original feature set in descending order, as it indicated in Fig.4(a), where (b) is a partly enlarged view of (a). Although there is some cross between extracted features, we still can see from figures that the singular values can generally differentiate different state of bearing failure in the initial phase of feature extraction.

3 Dimensionality reduction

3.1 Manifold

Manifold learning is a nonlinear dimensionality reduction method, which provides low-dimensional representations that are useful for processing and analysing data in a transformation-invariant way. And compared with previous nonlinear dimensionality reduction methods, manifold learning has powerful nonlinear dimensionality reduction and the ability to retain almost the same way as a simple linear dimension reduction algorithm [25]. The initial manifold learning mathematics is defined as follows: let Y∈R^d is a low-dimensional manifold, f:Y→R^d is a smooth embedding, where D>d. The data set {y_i } is randomly generated， and the data {x_i=f(y_i)} is mapped by f as the observed space. Indeed, manifold learning is to reconstruct f and {y_i } by given the sample set {x_i }. In other words, assuming that the data is sampled at a high-dimensional Euclidean space, manifold learning is to restore the low-dimensional manifold structure from high-dimensional data sampling in order to achieve data reduction and data visualization. In current studies, manifold learning algorithm including isometric feature mapping (IsoMap), locally linear embedding (LLE), Laplace eigenmaps (LE), local tangent space alignment (LTSA), etc.

3.1.1 Locally linear embedding (LLE)

The LLE algorithm attempts to preserve local order relation of the given data, which could be summarized as follows:

LLE: $X = \{x_1, x_2, \cdots, x_n\} \in R^d$ is the original dataset in the observation space; $Y = \{y_1, y_2, \cdots, y_n\} \in R^d (d \ll D)$is the low-dimensional manifold by a nonlinear mapping f(·). The purpose of LLE is to construct f(·) and Y by the given data X.

Step1: Find K nearest neighbours of each vector X_i;

Suppose the data consist of N vectors, choose K nearest neighbours measured by Euclidean distance.

Step2: Compute the weights and reconstruct;

Reconstruction errors are computed by the function below:

$$\varepsilon(W) = \sum_{i=1}^{N} |x_i - \sum_j w_{ij} x_i|^2 \quad (3)$$

Where w_{ij} represent the importance between the jth data point and the ith neighbor point. Assuming that each data point is reconstructed by its K neighbors, if x_j does not belong to the set, there is $w_{ij} = 0$; and $\sum_j w_{ij} = 1$, then

$$\begin{aligned}\varepsilon(W) &= \sum_{i=1}^{N} |x_i - \sum_j w_{ij} x_j|^2 \\ &= \sum_{i=1}^{N} |\sum_{j=0}^{k} w_{ij}(x_i - x_{ij})|^2 = \sum_{i=1}^{N} |(x_i - x_{ij}) w_i|^2 \\ &= \sum_{i=1}^{N} ((x_i - x_{ij}) w_i)^T ((x_i - x_{ij}) w_i) \\ &= \sum_{i=1}^{N} (w_i)^T (x_i - x_{ij})^T (x_i - x_{ij}) w_i \\ &= \sum_{i=1}^{N} (w_i)^T C_{jk} w_i\end{aligned} \quad (4)$$

By using the Lagrange multiplier, we can get

$$L(w) = \sum_{i=1}^{N} (w_i)^T C_{jk} w_i + \lambda(\sum_{j=1}^{k} w_{ij} - 1) \quad (5)$$

$$(\partial L) / (\partial w_i) = C_{jk} w_i + \lambda \quad (6)$$

So the weights are defined by

$$w_j = \frac{\sum_k C_{jk}^{-1}}{\sum_{lm} C_{lm}^{-1}} \quad (7)$$

Step3: Compute output vectors;

Y_i is best reconstructed by the weights W_{ij}, the embedding cost function is defined as follows:

$$\varepsilon(Y) = \sum_{i=1}^{N} |Y_i - \sum_j w_{ij} Y_j|^2 \quad (8)$$

Where $\sum_{i=1}^{N} Y_i = 0, \frac{1}{N} \sum_{i=1}^{N} Y_i Y_i^T = I$, and I is an $d \times d$ unit matrix. Minimizing the function above:

$$\begin{aligned}\min \varepsilon(Y) &= \sum_{i=1}^{N} |Y_i - \sum_j w_{ij} Y_j|^2 = \sum_{i=1}^{N} |YI_i - YW_i|^2 \\ &= \sum_{i=1}^{N} |Y(I_i - W_i)|^2 = |Y(I - W)|^2 = Y(I - W)(I - W)^T Y^T \\ &= YMY^T\end{aligned} \quad (9)$$

Similarly, by using the Lagrange multiplier, we can get

$$L(Y) = YMY^T + \lambda(NI - YY^T) \quad (10)$$

$$(\partial L) / (\partial Y) = MY^T + \lambda Y^T \quad (11)$$

$$MY^T = \lambda Y^T \quad (12)$$

Therefore we can get the bottom d+1 eigenvectors of M as the optimal d-dimensional embedding vectors [26].

3.1.2 Supervised locally linear embedding (SLLE)

LLE has good performance of discovering the intrinsic structure of nonlinear high-dimensional data and it is helpful to reduction dimensionality analysis. However, LLE cannot meet the needs of some certain aspects. To be precise, LLE is unsupervised learning mode, and not taking the class information of the data into account.

SLLE [27] is presented on the basis of LLE, in the processing of the first step increases the category information of the sample point. When calculating the distance between point and point, SLLE using the following equation:

$$D_{new} = D + \alpha \max(D) \cdot \Delta \quad (13)$$

Where D denotes the Euclidean distance between the sample points; max(D) represents the maximum distance between class and class; Δ is 0 or 1, when two points belong to the same class, $\Delta = 0$, else $\Delta = 1$; α is the experience parameter, and $\alpha \in [0,1]$, when $\alpha = 0$, SLLE is the same to LLE [28].

In virtue of taking full advantage of the information of labelled data and local neighbour structure, SLLE can obtain the whole intrinsic geometry of the dataset, and has good performance of data classification.

3.1.3 R-NSLLE

SLLE algorithm has a good pattern recognition peculiarity; it has been used in many fields. However, SLLE still exists some defects. K is the key coefficient to SLLE, and it cannot be chosen too small or too big otherwise the output vectors are not similar to the original structure. For choosing the best value of K, this paper put forward a new supervised locally linear embedding, that is R-NSLLE as it can be seen in Fig.5.

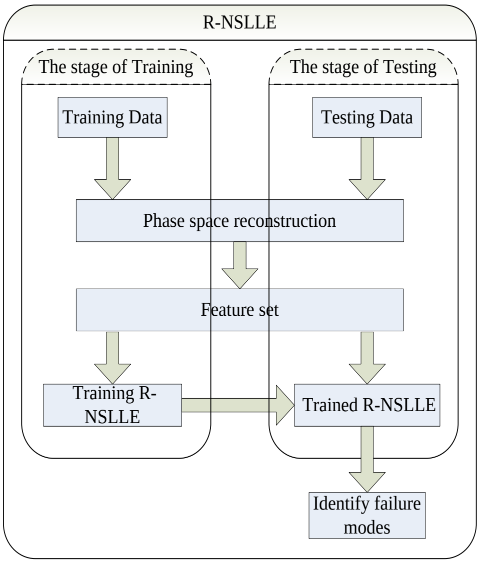


Fig.5 R-NSLLE

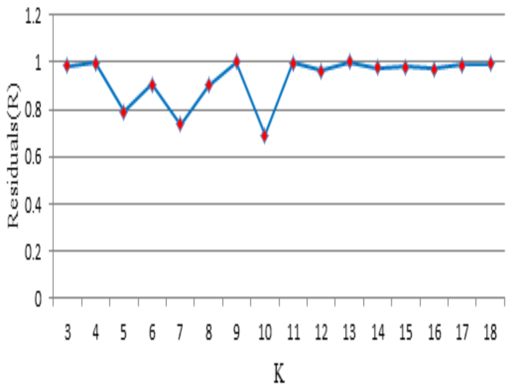


Fig.6 Residual of Euclidean distance

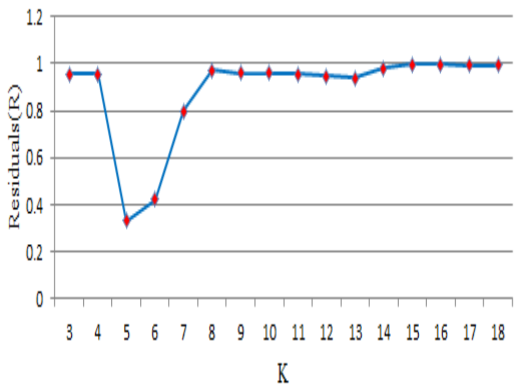


Fig.7 Residual of Geodesic distance

(1) Residual

Residual is defined as follows:

$$\text{Residual} = 1 - Corr^2(D_x - D_y) \quad (14)$$

Where $\text{Corr}(D_x - D_y)$ is standard linear correlation coefficient, D_x is the Euclidean distance matrix of original high-dimensional space, and D_y is the Euclidean distance matrix of low-dimensional embedding space. Residual can reflect the extent of keeping the original distance after dimensionality reduction, the more it is close to zero indicates that the effect of data embedded is better. The original fault feature set of 100×17 is used for test. For different values of K, Residual is calculated, as it is shown in Fig.6.

As can be seen from Fig.8, we can get k=10 is the best result. However, we know in advance that when k=10 it has poor clustering effect. Considering the nature structure of manifold, we introduce the concept of geodesic distance [29]. The following Fig.9 llustrates the concept of geodesic distance. A and B are any two points on the manifold, a straight line connecting between A and B is frequently used Euclidean distance. But the Euclidean distance can’t reflect the true relationship between A and B. On the manifold one of the shortest curve through A and B is called geodesic distance. For the equation of Residual we use geodesic distance instead of Euclidean distance and get the result in Fig.7. Different from the former, K=5 is the best choice, which is exactly the result what we want. In the subsequent analysis, we can still see the correctness of the results.

(2) The evaluation of residual

In order to further verify the validity of the method of residual, we here use CV to evaluate. CV is the coefficient of variation, which can describe the degree of aggregation in each fault state. For different values of K, CV is calculated, as it shown in Fig.10 From the Fig.10 we can see that only when k=5 that the value of CV is nearly to zero. In another words, when k=5, almost every fault type has a high cohesion, and the output vectors can maintain the original structure. That is the results are the same as the method of Residual. So we can come to the conclusion that Residual is a good method to choose the best value of K.

Besides, nonlinear supervised distance is used instead of linear supervised distance to improve the ability of the supervision of samples. The detail algorithm of R-NSLLE can be described as follow steps:

R-NSLLE: $X = \{x_1, x_2, \cdots, x_n\} \in R^d$ is the original dataset in the observation space;

$Y = \{y_1, y_2, \cdots, y_n\} \in R^d (d \ll D)$ is the low-dimensional manifold by a nonlinear mapping f(·).

Step1: Find K nearest neighbours of each vector X_i;

For choosing the optimal value of K, we use the method of Residual,

$$\text{Residual} = 1 - Corr^2(D_x - D_y) \quad (15)$$

Where $\text{Corr}(D_x - D_y)$ is standard linear correlation coefficient, D_x is the geodesic distance matrix of original high-dimensional space, and D_y is the Euclidean distance matrix of low-dimensional embedding space.

We use nonlinear supervised distance to choose K nearest neighbours：

$$D_{new}' = \begin{cases} \sqrt{e^{\frac{D^2}{\alpha}}}, & L_i \neq L_j \\ \sqrt{1 - e^{\frac{-D^2}{\beta}}}, & L_i = L_j \end{cases} \quad (16)$$

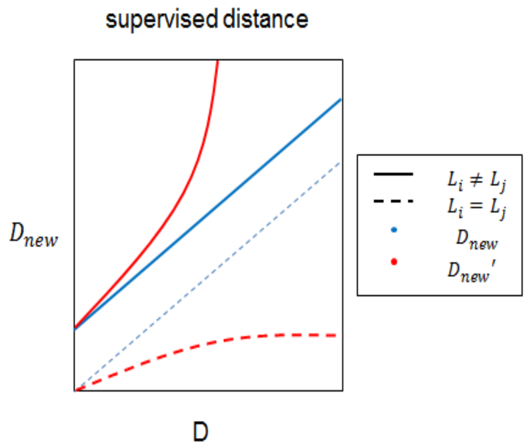


Fig.9 New nonlinear supervised distance

Where D is the distance of original data set, if $L_i = L_j$ it means the point i and j belong to the same class, otherwise they belong to different classes. α and β are adjustable parameters, which can be adjusted according to the degree of sparse data.

It can be seen from Fig.8 that with the growth of D the growth rate of between-class is far greater than the growth rate of within-class. That is this nonlinear supervised distance can enhance the discriminant ability of the algorithm.

Step2: Compute the weights and reconstruct;

Step3: Compute output vectors.

Step2 and step3 are the same to the steps of LLE.

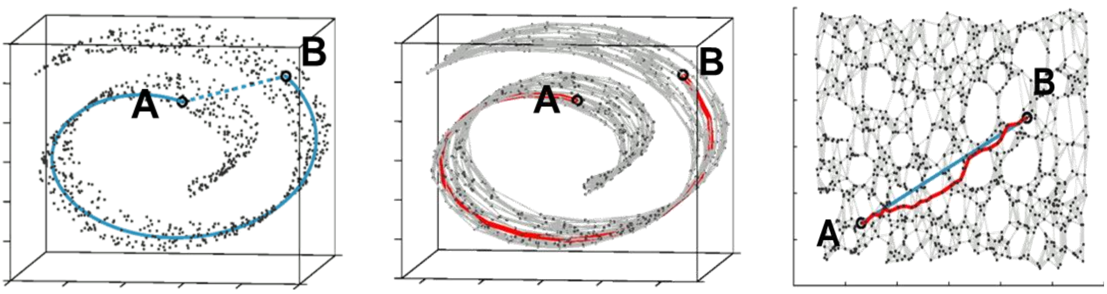


Fig.10 Geodesic distance

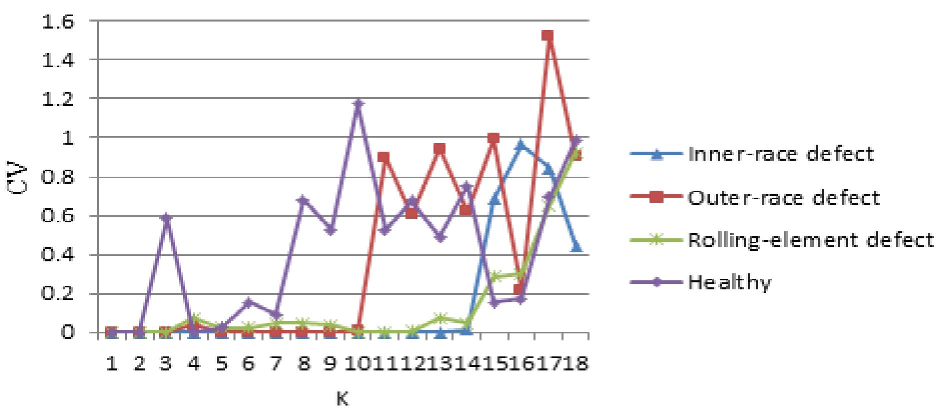


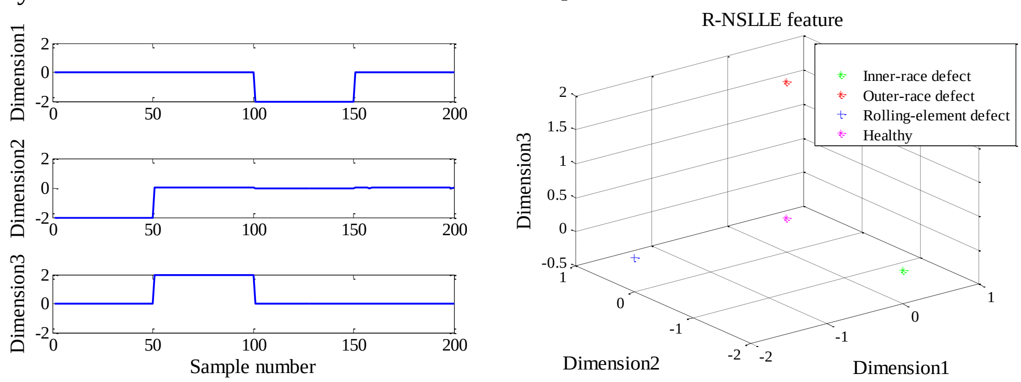
Fig.11 CV value

3.2 Experimental analysis

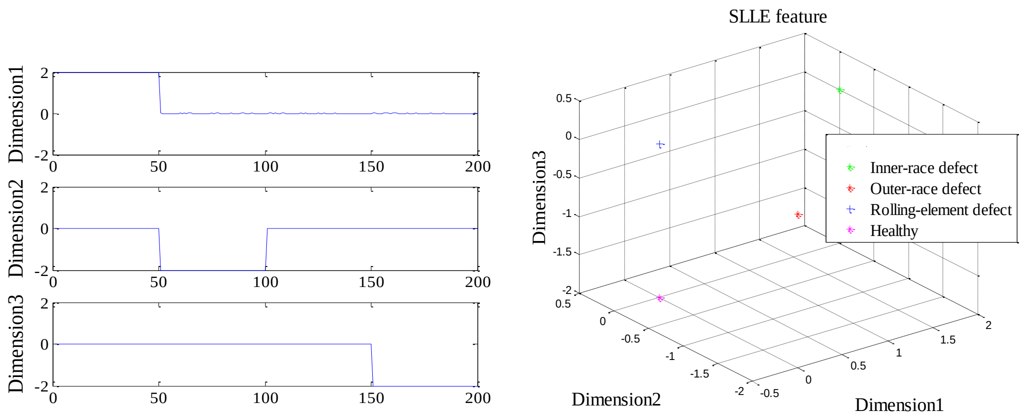
The test data acquired from experiment is weak failure data of rolling bearings. In the diagnosis of future research, we can construct computer and communication systems for data collecting and analysing [30]. Here the experimental platform is simply connected to the computer. The fault size is 0.5mm× 0.15mm with the sampling frequency 100 kHz. The experimental data obtained by sensor is one-dimensional vibration signal. We randomly selected 200 sets of sample data, and get the feature set of 200×17 after singular value decomposition. R-NSLLE is used for feature dimensionality reduction, which is compared with another nonlinear method called KPCA and other manifold learning SLLE, LLE and LTSA. We can get the data set of 200×3 after dimensionality reduction. (We know each method has its own optimal reduced dimensionality, experiments show that 3 is more appropriate).

Analyse the data with different fault types. As the parameter d=3, the calculated first three features are shown in the left of Fig.11. In the figure, the first 50 samples are inner-race defect data, 50-100 are outer-race defect data, 101-150 are rolling-element defect data, and 151-200 are the normal state data. The dimensionality reduction results of features for all the samples are given in the right of Fig.11 and each coloured dot represents a sample. As seen from the figure, the green part represents the inner-race defect, the red section represents the outer-race defect, the blue part represents the rolling-element defect, and the pink part represents the normal state.

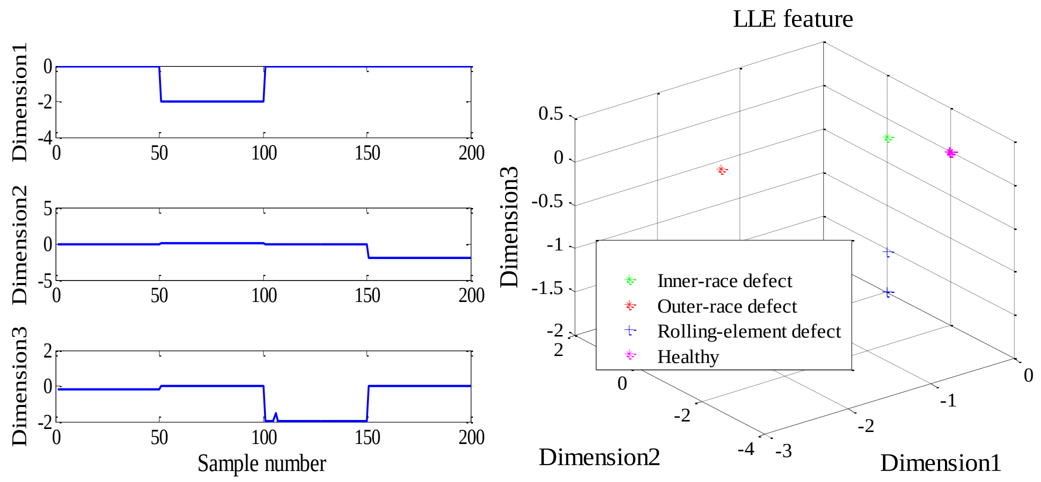
As we can see from figures above, for weak fault data of rolling bearings R-NSLLE has a better result in comparison with SLLE, LLE, LTSA and KPCA. As the results shown in KPCA feature, for a given fault it is not very stable parameters of each dimension. That is KPCA is susceptible to noise and other external conditions, it will affect the fault recognition. LTSA does not recognize the weak fault; it is associated with the defect itself. To some extent, LLE can recognize the weak fault, but there are still some external interference conditions. It can be seen that SLLE also has a good effect in spite of there are some fluctuations in Dimension1. But we must be clear that it was tried many times to choose the best value of K in the first step of SLLE. However, R-NSLLE is more robust via utilizing class information to guide the procedure of nonlinear mapping. R-NSLLE enhances local within-class relations and helps to classification. To further compare the effectiveness of the algorithm, the analysis of quantitative evaluation is discussed next.



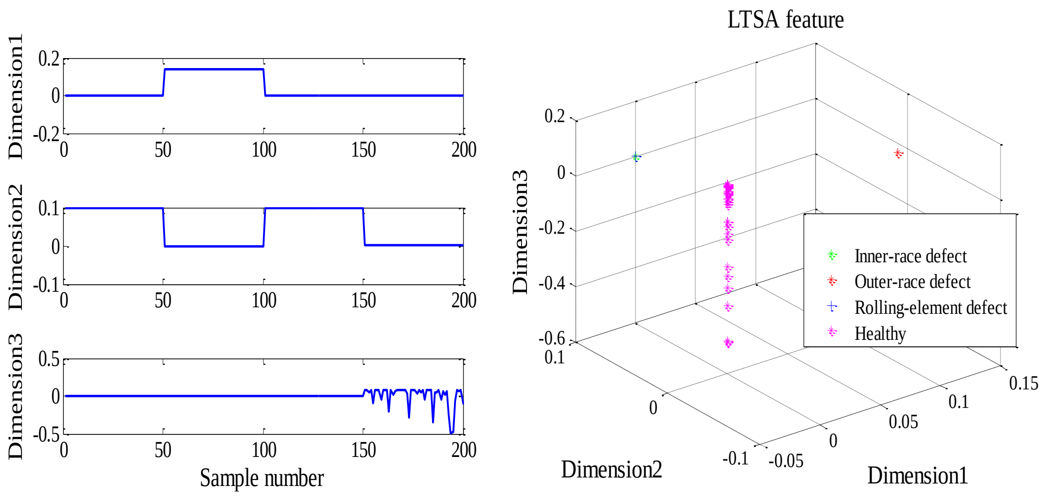
(a) R-NSLLE feature



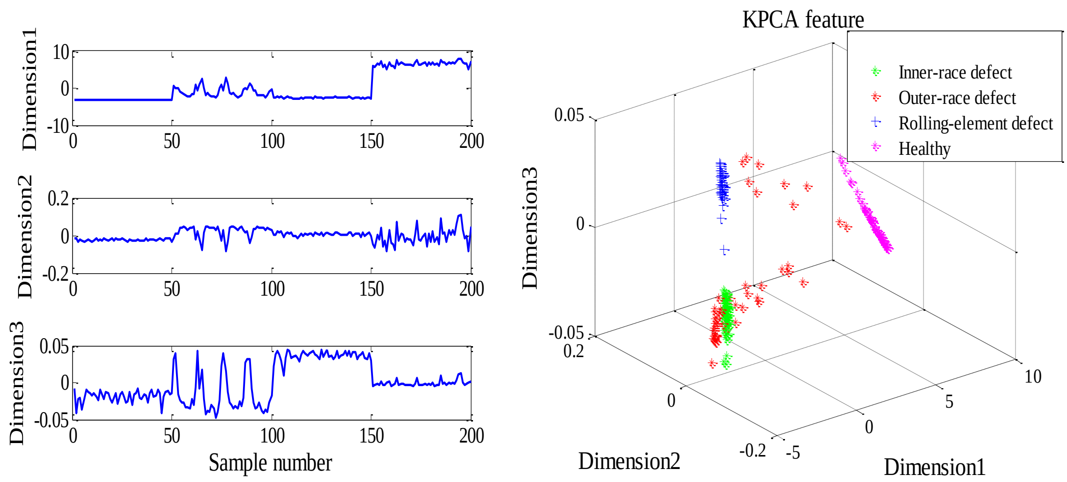
(b) SLLE feature



(c) LLE feature



(d) LTSA feature



(e) KPCA feature

Fig.12 Rolling features by dimensionality reduction

4 Quantitative evaluation

4.1 Accuracy

ELM is a new method based on single-hidden layer feedforward neural network (SLFN), which has a better capacity of generalization and a faster convergence speed [31, 32]. Nowadays, intelligent algorithm is widely used in machine learning and other fields, which can improve the effectiveness in practical applications [33, 34]. Here ELM is applied to fault identification.

After dimensionality reduction by R-NSLLE we can obtain characteristic parameter matrix of 200×3. Randomly generated training set and testing set, for each fault generated 30 samples as the training set, and the remaining 20 samples as the testing set. Thus the total training samples is 120 groups, as well as 80 groups of testing samples. These samples are then sent to ELM network. The results are indicated in Table 2. We also compare the Residual of each algorithm. Residual is a description of the data embedding, the smaller the value, the better it is.

It is obvious that R-NSLLE is better than other methods. As seen from the table 2, the diagnostic accuracy of R-NSLLE is up to 100%. Also, residual of R-NSLLE is the smallest in comparison with other algorithms. In other words, R-NSLLE is good to keep the structure of the original data after reducing dimension. It means that R-NSLLE method is more suitable for feature dimensionality reduction of rolling bearings. As for SLLE, the results are depends on the value of K. This will result in a higher complexity of the time. If the value of K is selected appropriate, the results are better. However inappropriate value of K will lead to poor results even down to 60%.

4.2 Silhouette coefficient

Silhouette coefficient [35] can be defined as follows, which is a measure used to describe the clustering effect. For the data set D, assume that D is divided into k clusters $C_1, C_2, \cdots, C_K, 0 \in C_I (1 \le i \le k)$, then

$$s(o) = \frac{b(o) - a(o)}{\max\{a(o), b(o)\}} \tag{17}$$

$$a(o) = \frac{\sum_{o,o' \in C_i, o \neq 0'} dist(o,o')}{|C_i| - 1} \tag{18}$$

$$b(o) = \min_{C_j: 1\le j \le k, j \ne i} \left\{ \frac{\sum_{o' \in C_j} dist(o,o')}{|C_j|} \right\} \tag{19}$$

Where $-1 \le s(o) \le 1$. $a(o)$ is a measure of compactness within one cluster, and $b(o)$ describes the dispersion between clusters. Thus $s(o)$ close to one means that the cluster contains o is compact and o is far away from other clusters. However, when the value of $s(o)$ is negative, we think that the clustering result is not acceptable. Here the average of $s(o)$ is used for evaluation, it can be the measure of how appropriately the data has been clustered [36].

The average of silhouette coefficient is calculated for R-NSLLE, SLLE, LLE, LTSA and KPCA features, respectively. As shown in Table 3, the average of silhouette is equal to one for every status of rolling bearings. It shows that R-NSLLE can fully identify all types of faults. SLLE and LLE also have a good result, but the silhouette coefficient of rolling-element defect is smaller, that can reduce the diagnostic accuracy. It is obvious that Silhouette coefficient is affected by the value of K. As for LTSA and KPCA, there exists a negative value of silhouette coefficient; it means that the result has a certain degree of error. So we can draw the conclusion that R-NSLLE can effectively extract the bearing features, and it has good capability of classification.

Table 2 Accuracy

Performance indicators	R-NSLLE	SLLE		LLE	LTSA	KPCA
Training sample	120	120		120	120	120
Test sample	80	80		80	80	80
Accuracy	100%	100%	60%	98%	86.25%	90%
Residual	0.3829	0.7932	0.9679	0.8982	0.6881	0.8218

Table 3 The average of silhouette coefficient

Silhouette coefficient	Inner-race defect	Outer-race defect	Rolling-element defect	Healthy
R-NSLLE	1	1	1	1
SLLE	1	1	0.9768	1
	1	-0.7906	1	-0.7061
LLE	1	1	0.8725	1
LTSA	0.9944	1	0.9592	-0.3799
KPCA	0.9755	-0.1527	0.7821	0.9862

5 Conclusions

Considering the complex mechanical state of rolling bearings, the general features of time-frequency characteristics are susceptible to noise, thus they do not fully reflect the operational status of rolling bearings. This paper proposes a new feature extraction method. Using the improved C-C algorithm, the optimal time delay and embedding dimension are obtained from the one-dimensional vibration signal of a rolling bearing. This signal contains a wealth of information after the phase space is reconstructed. Then, singular values are extracted from the high-dimensional signal, which are used as fault features. Following this, R-NSLLE is used for nonlinear dimensionality reduction. Experimental results show that R-NSLLE can better mine the intrinsic low-dimensional data from high-dimensional data, which has certain advantages in terms of the weak fault diagnosis of rolling bearings.

Acknowledgment

This project is supported by National Natural Science Foundation of China (Grant No. 51375037, 51135001) and Program for New Century Excellent Talents in University (NCET-12-0759).

References:

[1] P. Wang, H. F. Yuan, H. Q. Wang, X. Cao and X. W. Wang, Rolling Element Bearing Fault Diagnosis Based on Symptom Parameter Wave of Acoustic Emission Signal, *Advanced Science Letters*, Vol. 13, No. 1, 2012, pp. 667-670.

[2] M. Cococcioni, B. Lazzerini and S. L. Volp, Robust Diagnosis of Rolling Element Bearings Based on Classification Techniques, *IEEE Transactions on Industrial Informatics*, Vol. 9, No. 4, 2013, pp. 2256-2263.

[3] H. Qiao, P. Zhang, and D. Wang, An Explicit Nonlinear Mapping for Manifold Learning, *IEEE Transactions on Cybernetics*, Vol. 43, No.1, 2013, pp. 51-63.

[4] G. W. Wang, J. Zhuang and D. H. Yu, An Investigation of Applying Manifold Learning to Diagnose Machinery Faults, *Journal of Engineering Mathematics*, Vol. 29, No. 4, 2012, pp. 593-599.

[5] W. Peng, H. J. Wang, and X. L. Xu, Fault Diagnosis Model Based on Local Tangent Space Alignment and Support Vector Machine, *Chinese Journal of Scientific Instrument*, Vol. 33, No.12, 2012, pp. 2789-2795.

[6] Z. X. Li, X. P. Yan, etc, A New Data Mining Approach for Gear Crack Level Identification Based on Manifold Learning, *Mechanika*, Vol. 18, No.1, 2012, pp. 29-34.

[7] H. B. Xiao, and X. P. Xie, Feature Extraction of Diesel Engine Wear Fault Based on Local Tangent Space Alignment, *International Conference on Structural Engineering, Vibration and Aerospace Engineering*, Vol. 482, 2014, pp. 179-182.

[8] S. Roweis and L. Saul, Nonlinear Dimensionality Reduction by Locally Linear Embedding, *Science*, vol. 290, No. 5500, 2000, pp. 2323-2326.

[9] Q. S Jiang, J. Y. Lv, etc, New Approach of Intelligent Fault Diagnosis Based on LLE Algorithm, *Chinese Control and Decision Conference*, 2008, pp. 522-526.

[10] S. M. Xiang, F. P. Nie, ect, Regression Reformulations of LLE and LTSA with Locally Linear Transformation, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 41, No.5, 2011, pp. 1250-1262.

[11] M. Li, J. Xu, J. Yang, D. Yang, and D. Wang, Multiple Manifolds Analysis and its Application to Fault Diagnosis, *Mech. Syst. Signal. Process*, Vol. 23, 2009, pp. 2500–2509.

[12] G. Wang, Y. Liu, and X. Zhao, Fault Diagnosis of Rolling Bearings Based on LLE-KFDA, *Mater. Sci. Forum*, Vol. 626, 2009, pp. 529–534.

[13] B. W. Li and Y. Zhang, Supervised Locally Linear Embedding Projection (SLLEP) for Machinery Fault Diagnosis, *Mechanical Systems and Signal Processing*, Vol. 25, 2011, pp. 3125–3134.

[14] M. A. Sanjith, B. Krishnakumar and S.A.V. Satya Murty, Roller Element Bearing Fault Diagnosis using Singular Spectrum Analysis, *Mechanical Systems and Signal Processing*, Vol. 35, 2013, pp. 150–166.

[15] X.Y. Jin, L. Lin, S. S Zhong and G. Ding, Rotor Fault Analysis of Classification Accuracy Optimition Base on Kernel Principal Component Analysis and SVM, *Procedia Engineering*, Vol. 15, 2011, pp. 5279 – 5283.

[16] H. F. Yuan, C. Ji, and H. Q. Wang. Intelligent Fault Diagnosis Based on KPCA-RS of Rolling Bearing, *8th WCEAM & 3rd ICUMAS*, 2013.

[17] Z. Y. Zhang, H. Y. Zha, Principal Manifolds and Nonlinear Dimensionality Reduction via Tangent Space Alignment, *Journal of Shanghai University*, Vol. 8, No. 11, 2004, pp. 406-424.

[18] H. F. Yuan, F. M. H. Q. Wang, Using Evaluation and Leading Mechanism to Optimize Fault Diagnosis Based on Ant Algorithm, *Energy Proscenia*, Vol. 1, No. 6, 2012, pp. 112-116.

[19] H. S. Kim, R. Eykhol and J. D. Salas, Nonlinear Dynamics, Delay Times, and Embedding Windows, *Physica D*, Vol. 127, No.1-2, 1999, pp. 48-60.

[20] Z. B. Lu, Z. M. Cai, etc, Determination of Embedding Parameters for Phase Space Reconstruction Based on Improved C-C Method, *Journal of System Simulation*, Vol. 19, No.11, 2007, pp. 2527-2538.

[21] Y.Q. Qin, W. D. Cai, etc, Research on Phase Space Reconstruction of Nonlinear Time Series, *Journal of System Simulation*, Vol. 20, No.11, 2008, pp. 969-2973.

[22] Q. B. He, Time–frequency Manifold for Nonlinear Feature Extraction in Machinery Fault Diagnosis, *Mechanical Systems and Signal Processing*, Vol. 35, 2013, pp. 200–218.

[23] B. Kilundu, X. Chiementin and P. Dehombreux, Singular Spectrum Analysis for Bearing Defect Detection, *Vib. Acoust*, Vol. 133, No.5, 2011, pp. 1-7.

[24] W. Z. Dong, Bearing Fault Diagnosis Method Based on EEMD and Difference Spectrum Theory of Singular Value, *Journal of Mechanical Strength*, Vol. 34, No. 2, 2012, pp. 183-189.

[25] X.Y Hou, Research on the Extraction Method Based on Manifold, Dalian University of Technology , 2009.

[26] S. L Zhao and S. A. Zhu, Face Recognition by LLE Dimensionality Reduction, I*nternational Conference on Intelligent Computation Technology and Automation*, 2011, pp. 121-123.

[27] Q. S. Jiang, J. Lu, and M. Jia, New Method of Fault Feature Extraction Based on Supervised LLE, *Proceedings of 2010 Chinese Control and Decision Conference*, 2010, pp. 1727 - 1731.

[28] S. Q. Zhang and L. M. Li, Speech Emotion Recognition Based on an Improved Supervised Manifold Learning Algorithm, *Journal of Electronics and Information Technology*, Vol. 32, No.11, 2010, pp. 2724-2729.

[29] J. B. Tenenbaum, V. D. Silva and J. C. Langford, A Global Geometric Framework for Nonlinear Dimensionality Reduction, *Science*, Vol. 290, No.5500, 2000, pp. 2319-2323.

[30] S. Staines A., Neri F., A Matrix Transition Oriented Net for Modeling Distributed Complex Computer and Communication Systems, *WSEAS Transactions on Systems*, WSEAS Press (Athens, Greece), Vol. 13, 2014, pp. 12-22.

[31] G. B. Huang, Q. Y. Zhu, and C. K. Siew, Extreme Learning Machine: Theory and Applications, *Neurocomputing*, Vol. 70, 2006, pp. 489–501.

[32] G. B. Huang, D. H. Wang and Y. Lan, Extreme Learning Machines: a Survey, Vol. 2, 2011, pp. 107-122.

[33] M Papoutsidakis, D Piromalis, F Neri, M Camilleri, Intelligent Algorithms Based on Data Processing for Modular Robotic Vehicles Control, *WSEAS Transactions on Systems*, *WSEAS Press (Athens, Greece)*, Vol. 13, 2014, pp. 242-251.

[34] Camilleri M., Neri F., Papoutsidakis M, An Algorithmic Approach to Parameter Selection in Machine Learning using Meta-Optimization Techniques, *WSEAS Transactions on Systems*, *WSEAS Press (Athens, Greece)*, Vol. 13, 2014, pp. 202-213.

[35] Y. C. Liu and X. D. Gao, Ensembling Clustering Validation Indices, *Computer Engineering and Applications*, Vol. 47, No.19, 2011, pp. 15-17.

[36] J. W. Han and M. Kamber, *Data Mining: Concepts and Techniques*, China Machine Press, 2012.