Multi-Layer Kernel Learning Method Faced on Roller Bearing Fault Diagnosis

Guangbin Wang
Engineering Research Center of Advanced Mine Equipment, Ministry of Education,
Hunan University of Science and Technology, China
Email:jxxwgb@126.com

Yilin He and Kuanfang He
Hunan Provincial Key Laboratory of Health Maintenance for Mechanical Equipment,
Hunan University of Science and Technology, China
Email:281576286@qq.com, hkf791113@163.com

***Abstract*—Bearing fault is the major fault of the rotating machinery, in order to better identify the fault of bearing, the multi-layer kernel learning methods based on local tangent space alignment (LTSA) and support vector machine (SVM) are proposed. In this method, the supervised learning is embedded into the improved local tangent space alignment algorithm, realize fault feature extraction and new data processing for equipment fault signal, and then correctly classify the faults by non-linear support vector machine. The experiment result for roller bearing fault diagnosis shows that SILTSA-SVM method has better diagnosis effect to related methods.**

***Index Terms*—fault diagnosis, multi-layer kernel, SVM, supervised, LLTSA, LLTSA**

I. Introduction

The basic problem of fault diagnosis is to obtain fault status by extracting feature, designing decision functions based on the status information of equipment. Feature extraction and pattern recognition is the core of the problem. Bearing fault is the major fault of the rotating machinery, the basic fault classes have inner ring fault, outer ring fault, ball fault and so on.

With the application of the multiple sensors monitor technology, the status information of the equipment sampled by multiple sensors become more and more. On one hand, a lot of data can make us learn more about the equipment operation, on the other hand, it is difficult for these data to effectively deal with, and eventually may cause the problem of the dimensionality disaster.

In 2000, S. Roweis and H.S. Seung simultaneously had published the research papers about the manifold learning in Science [1-2], proposed Isometric feature Mapping (ISOMAP) [1] algorithm and Locally Linear Embedding (LLE) [2] algorithm, and successfully applied them to the graph and characters recognition. As a starting point, the researchers had launched a variety of algorithms, such as Laplace feature Mapping (LE) [3], Local Tangent Space Alignment (LTSA) [4] and so on.

Manifold learning can be used to solve dimensionality disaster problem, but they cannot process incremental data. In order to solve the problem, some scholars had proposed linear methods of manifold learning were, such as Local Preserving Projection (LPP) [5], Neighborhood Preserving Embedding (NPE) [6], Local Linear Tangent Space Alignment (LLTSA) [7]. Other scholars had proposed incremental algorithms, such as incremental ISOMAP [8], incremental LLE [9] and LTSA [10].

All manifold learning algorithms and their improved algorithms are also considered as generalized kernel methods under data correlation [15]. When these algorithms were applied to equipment fault diagnosis, the supervision information of fault signals were not fully used, so the effect of diagnosis is limited. In this paper, the supervised learning and multi-layer kernel idea were introduced to improved LTSA, SLLTSA-SVM and SILTSA-SVM algorithms are proposed. Two methods may realize feature extraction and pattern recognition of the fault signal by supervised learning. At the same time new data are processed by linear or incremental method, and fault type is correctly identified by nonlinear SVM. The experiment result for roller bearing fault diagnosis shows that supervised multi-layer kernel method had better identified effect than other related methods.

II. Manifold Learning and Kernel Method

Manifold learning and kernel method are machine learning methods to acquire useful information by mapping; many researchers hope to unify manifold learning to kernel method’s framework. Based on kernel idea, Choi had proposed kenerl isomap method [11], Wang had proposed kenerl orthogonal local fisher discriminate analysis [12], Weinberger had obtained the relation between KPCA and manifold method by semi-definite programming [13] , Yan had unified manifold learning to map analysis’s framework [14]. Huang had deep studied manifold learning such as ISOMAP, LLE and LE, they believe these classical methods can be seen as KPCA under data correlation from the perspective of kernel technology [15], the Mercer theorem is satisfied in these methods by respectively the means of constructing distance matrix, laplacian matrix and the weight matrix.

Manifold learning method have the similar idea and some common characteristics, firstly, local neighborhood structures of the sample points on manifold is constructed,

doi:10.4304/jsw.7.7.1531-1538

then we use these local neighborhood structures to map sample points to global low dimensional space. The main difference lie in the local neighborhood structures and embedded method by using local neighborhood structures to construct the global low dimension embedded coordinates. In LLE, global coordinate is built by mining errors between the local coordinate and global coordinate, instead of this method, LTSA implements the alignment from local model to global coordinate by affine transformation. Therefore, LTSA is the improvement of LLE, which can also be seen as KPCA under data correlation.

SVM is a kind of the typical kernel method proposed by Vapnik based on the largest interval hyperplane, Mercer kernel, and convex quadratic programming and relaxation techniques [16]. SVM develops from the optimal classification surface in linear separable condition. The optimal separating surface is the surface that has correct classification effect, and also the largest interval between different categories.

Considering samples $(X_i, y_i)(i = 1,2,\cdots,n)$, category label is $X \in R^d, y \in \{-1,-1\}$, linear discriminate function is

$$g(x) = W \bullet X + b \quad (1)$$

In linear inseparable condition, the samples in input space are mapped to high dimensional space through the kernel function meeting consistent with the Mercer conditions, and we can obtain optimal classification surface. This surface is described by (2)

$$\left.\begin{aligned} &\min \phi(W) = \frac{1}{2}\| W \|^2 = \frac{1}{2}(W \bullet W) \\ &s.t \quad y_i[(W \bullet X_i) + b] - 1 \geq 0 \quad (i = 1,2\cdots n) \end{aligned}\right\} \quad (2)$$

Making use of Lagrange algorithm, the dual optimization equation is obtained as (3)

$$\left.\begin{aligned} &\min Q(\alpha) = \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(X_i, X_j) - \sum_{i=1}^{n}\alpha_i \\ &s.t \quad \alpha_i \geq 0 \quad (i = 1,2\cdots n) \\ &\sum_{i=1}^{n} y_i \alpha_i = 0 \end{aligned}\right\} \quad (3)$$

Where $\alpha_i (i = 1,2\cdots n)$ is Lagrange operator, $\alpha_i \in [0, C]$, C is slack variable, the optimal classification function is as followed

$$f(x) = \operatorname{sgn}\left\{ \sum_{i=1}^{n} \alpha_i^* y_i K(X_i, X) + b^* \right\} \quad (4)$$

(3) is SVM model, b^* is the solution of $y_i[\sum_{i=1}^{n} \alpha_i^* y_i K(X_i, X) + b^*] = 1$.

The steps of SVM are, appropriate kernel function is firstly selected, the optimum equation is solved and obtain support vector and the corresponding Lagrange operator, finally obtain the optimum classification function by which classify correctly.

III. Two improved Algorithms based on LTSA Incremental local tangent space alignment

A. Incremental local tangent space alignment (ILTSA)

The following is the main theory of ILTSA. In LTSA, we can obtain matrix S_i with right singular feature vectors corresponding to the maximum m feature values by making singular value decomposition to $X_i - \bar{x}_i 1_k^T$,where $\bar{x}_i = \frac{1}{k}\sum_{j=1}^{k} x_{ij}$ .Local tangent space coordinate Θ_i is given by projection of data on base Q_i ,where

$$\Theta_i = Q_i^T (X_i - \bar{x}_i 1_k^T) = [\vartheta_1^i, \theta_2^i \cdots \theta_k^i] \quad (5)$$

Let affine transformation matrix $L_i \in R^{d\times d}$ and low dimensional coordinates $Y = [Y_1, Y_2, \cdots Y_n]$, in this matrix, $Y_i = [y_{i1}, y_{i2}, \cdots y_{ik}]$. Y is obtained by minimizing the reconstruction error E_i, Optimization problem can be described by (6)

$$\min_{L_i Y_i} \sum_i \| E_i \|_2^2 = \min_{L_i Y_i} \sum_i \| Y_i (I - \frac{1}{k_i} ee^T) - L_i \Theta_i \|_2^2$$

$$\min_Y \| YSW \| = \min_Y tr(YSWW^T S^T Y^T) \quad (6)$$

Where, $YS_i = Y_i$, $W = diag(W_1, \cdots W_i, \cdots W_N)$, $W_i = (I - \frac{1}{k} ee^T)(I - \Theta_i^+ \Theta_i)$.

In constraints of $YY^T = I$, the optimal m dimensional coordinates Y is constituted of eigenvectors corresponding to the m minimum eigenvalues of $B = SWW^T S^T$. There, the low dimensional martrix $Y = [u_2, u_3, \cdots u_{m+1}]^T$. To solve the singular problem of B, firstly data are mapped to the subspace by PCA, then the main feature are extracted by LTSA. In ILTSA, the problem to be solved is how to obtain the global low-dimensional coordinate τ_{n+1} of the new test point x_{n+1}.

Let X_A be matrix including x_{n+1} , the solution process to τ_{n+1} is equivalent to minimizing the local reconstruction error ε_{n+1}^i

$$\begin{aligned} \| \varepsilon_{n+1}^i \|_2^2 &= \| \tau_{n+1}^i - [\bar{\tau}_i + L_i \theta_{n+1}^i] \|_2^2 \\ &= \| \tau_{n+1}^i - [\bar{\tau}_i + T_i \theta_{n+1}^i] \|_2^2 \end{aligned} \quad (7)$$

Optimization problem may be expressed by the following as (8)

$$\min_{\tau_{n+1}} \| \varepsilon_{n+1}^i \|_2^2 = \min_{\tau_{n+1}} \| \tau_{n+1}^i - [\bar{\tau}_i + T_i \Theta_i^+ \theta_{n+1}^i] \|_2^2 \quad (8)$$

In fact, the process of calculating global low dimensional solution is also the process of eigenvalue decomposition to updated B_{new}. The update process is as follows, let $B_{new}(I_i, I_i) = 0$, $I_i = \{i_1, i_2, \cdots i_k\}$ be neighbors subscript of x_i for $x_i \in X_A$, and then B_{new} is updated by the following (9)

$$B_{new}(I_i, I_i) \leftarrow B(I_i, I_i) + I - G_i G_i^T \quad (9)$$

In summary, SILTSA basic steps are as follows, firstly, the dimension of training data is reduced by manifold learning, secondly the new data point x_{new} is placed into the original data set , X_A and B_{new} are rebuilt , at last we can obtain τ_{new} of all new data point x_{new} by (8).

Linear local tangent space alignment (LLTSA)

LLTSA is a linear algorithm. Supposing the local tangent coordinates is Θ_i ,and then $\Theta_i = Q_i^T (X_i - \bar{x}_i 1_k^T) = [\vartheta_1^i, \theta_2^i \cdots \theta_k^i]$. According to LLTSA, (10) should come into existence

$$\arg\min_{x,\Theta,Q} \sum_{j=1}^{k} \| x_{ij} - (x + Q\theta_j) \|_2^2$$

$$= \arg\min_{\Theta,Q} \| X_i H_k - Q\Theta \|_2^2 \quad (10)$$

In (10), $H_k = I - ee^T / k$, Q is the orthogonal basis vector matrix in tangent space. Let affine transformation matrix $L_i \in R^{d \times d}$, and global low dimensional coordinates $Y = [Y_1, Y_2, \cdots Y_n]$, $Y_i = [y_{i1}, y_{i2}, \cdots y_{ik}]$. In LTSA, optimization problem can be described by (7) and (8), Among their formulas, $W = diag(W_1, \cdots W_i, \cdots W_N)$, $W_i = H_k (I - \Theta_i^+ \Theta_i)$, W_i can write

$$W_i = H_k (I - V_i V_i^T) \quad (11)$$

V_i is the right singular vector correspond to largest singular value of matrix $X_i H_k$. After adding constraint $YY^T = I_d$, Y will be unique. Supposing A is linear mapping, $Y = A^T X H_k$, the target function can be obtained by the following constrained optimization problem

$$\arg\min_{Y} tr(A^T X H_N B H_N X^T A)$$

$$A^T X H_N X^T A = I_d \quad (12)$$

In (12), $B = SWW^T S^T$. (12) can replaced by solving eigenvalue of (13).

$$X H_N B H_N X^T \alpha = \lambda X H_N X^T \alpha \quad (13)$$

Eigenvectors $\alpha_1, \alpha_2, \cdots \alpha_d$ correspond to eigenvalue $\lambda_1 < \lambda_2 < \cdots < \lambda_d$ constitute the linear mapping vector matrix $A_{SLLTSA} = [\alpha_1, \alpha_2, \cdots \alpha_d]$. In LLTSA, the singular problem of $X H_N X^T$ has been solved by PCA mapping before the algorithm.

Supervised ILTSA and Supervised LLTSA

ILTSA and LLTSA can effectively solve dimension reduction of the new data, but two algorithms could not take full advantage of the class information of data,so it is difficult to effectively extract the fault characteristics of mechanical equipment. The supervised idea is introduced to algorithm; it can better solve this problem.

Given $X = [x_1, x_2, \cdots x_N] \in R^d$, d is the data dimension, the euclid distance between any two points x_i and x_j is

$$D_{ij} = \sqrt{(x_i - x_j)(x_i - x_j)^T} \quad (14)$$

In LLSTA and ILTSA, the neighbor matrix of each point x_i is built by directly finding k points with minimum distance to x_i in D_{ij}. In Supervised algorithm, considering fault categories, the neighbors matrix is built by the following (15)

$$D'_{ij} = D_{ij} + \alpha \max(D_{ij}) \Delta \quad (15)$$

D'_{ij} is the generalized matrix after introducing category information. $\max(D_{ij})$ shows maximum distance between the class containing x_i and the class containing x_j. $\Delta = 0$ when x_i and x_j belong to the same category, otherwise $\Delta = 1$. $\alpha \in [0,1]$ is empirical parameter to control the distance between point sets.

LLTSA and ILTSA only maintains local geometry inside data while not consider category information when $\alpha = 0$, and maximizes the distance between categories while the capacity of local geometry's maintains will be weakened when $\alpha = 1$. We can obtain the generalized nearest neighborhood matrix $X_i = [x_{i1}, x_{i2}, \cdots x_{ik}]$ by finding k points with minimum distance to x_i .

In this paper, we introduce supervised idea to LLTSA and ILTSA; propose two supervised algorithms by using D'_{ij} instead of D_{ij} , SLLTSA and SILTSA.

The maximum likelihood estimation for intrinsic dimension

In LTSA and its improved algorithm, one key parameter directly affects calculation results; it is target dimension of analyzed data. The size of the main feature space is decided by target dimension, the choice of target dimension is stochastic and uncertain scope thus causes the efficiency of the dimension reduction lower.

Intrinsic dimension is also called topological dimension, it is required minimum number of independent parameters for data sets' description. Intrinsic dimension of signal determines distribution of phase points. Maximum Likelihood Estimation (MLE) is a new method of global intrinsic dimension estimation by establishing the likelihood function of distance between neighboring points.

Let x_i is Independent and identically distributed sample, y_i is the smooth embedded manifold in R^d , there has $x_i = g(y_i)$ .Assume $S_X(t)$ is a small sphere that x is core, t is radius, we can construct non-homogeneous binomial

$$N(t,x)=\sum_{i=1}^{n} I\{x_i \in S_X(t)\} \tag{16}$$

Similar to the process with poisson distribution, $\lambda(t)=f(x)V(d)dt^{d-1}$ is obtained, let $\theta=\log f(x)$, the likelihood function of observed process is

$$L(d,\theta)=\int_0^t \log\lambda(t)dV(t)-\int_0^t \lambda(t)dt \tag{17}$$

Let $\frac{\partial L}{\partial\theta}=0$, $\frac{\partial L}{\partial n}=0$, neighbor number k is regarded as the radius t of small ball, maximum likelihood estimation $\tilde{d}_k(x)$ of x is

$$\tilde{d}_k(x)=\left[\frac{1}{k}\sum_{j=1}^{k-1}\log\frac{T_k(x)}{T_j(x)}\right]^{-1} \tag{18}$$

Make x traverses the entire data set ,we can obtain intrinsic dimension d

$$d=\frac{1}{n}\sum_{i=1}^{n}\tilde{d}_k(x) \tag{18}$$

IV. Fault Diagnosis Model based on Multi-Layer Kernel Learning

Manifold learning and support vector machines are two machine learning methods on the basis of a kernel function, the former is an effective tool to solve non-linear feature extraction as the unsupervised learning algorithm, the latter as a monitor algorithm has strong classification ability, not only can reduce the computational complexity and storage space, but also can also reduce the generalization error. To make full of the advantages of these two methods respectively, we combine manifold learning and SVM, have proposed two multi-layer kernel learning model, ILTSA-SVM and ILTSA-SVM.

Supposing sampling frequency is fs HZ, sample vector x_i is made up of data sampling by rotating machinery in one cycle, the rotate speed is v HZ, so data each sample x_i have $d=fs/v$ points which d is dimension. Let $X=[x_1,x_2,\cdots x_n]\in R^d$ be training sample matrix, $Y=[y_1,y_2,\cdots y_p]\in R^d$ be test sample matrix, $m<<d$ is the dimension of feature space. The basic steps of SILTSA-SVM fault diagnosis can be described as followed

Step 1.The intrinsic dimension of the data is estimated by means of MLE, and then the choice of target dimension based on this dimension.

Step 2. Data extraction and processing based on PCA projection.

Step 3. Select the number of the neighborhood k, calculate the feature matrix of training samples $Z^x=[\tau_1,\tau_2,\cdots\tau_n]\in R^m$.

Step 4. Calculate the low dimensional embedding of each new data $y_i\in R^d, i=1,2\cdots p$ in $Y=[y_1,y_2,\cdots y_p]$ individually; obtain the feature matrix of test samples $Z^Y=[\tau_1^y,\tau_2^y,\cdots\tau_p^y]\in R^m$.

Step 5. Using the feature matrix of training samples Z^x, obtain optimum classification function $f(x)$ and make fault diagnosis to test data by the SVM .

The basic steps of SLLTSA-SVM fault diagnosis can be described as followed

Step 1. The intrinsic dimension of the data is estimated by means of MLE, and then the choice of target dimension based on this dimension.

Step 2. Data extraction and processing based on PCA projection.

Step 3. Select the number of the neighborhood k. With SLLTSA we can obtain linear mapping matrix $A=A_{PCA}A_{LLTSA}$.

Step 4. Calculate the matrix $Z^x=A^TXH_N\in R^m$ of training samples, the matrix $Z^y=A^TYH_N\in R^m$ of test samples .

Step 5. Choose kernel function and corresponding kernel parameter, penalty factor C, construct nonlinear SVM model. Using the feature matrix Z^x of training data, optimizing SVM model, and obtain optimal classification decision function.

Step 6. According to the SVM model, gain the fault type of feature matrix Z^y of test date.

V. Multi-Layer Kernel Learning experiment on Roller Bearing Fault Diagnosis

Rolling bearing is important supporting and easily damaged parts in the mechanical system. According to incomplete statistics, about 30 percent of rotating machinery fault is caused by bearings. In this experiment, bearing vibration acceleration normal and fault sample data come from the CWRU bearing data center site [17]. Figure 1 is the rolling bearing test experiment stand.

Bearing damage is artificially produced by using electric discharge machines. Acceleration sensors were placed in base, driver-side and fan-side of the test motor. Bearing is the deep groove ball bearing for SKF6205, and its inner and outer diameter are 25mm and 52mm respectively, fundamental frequency of bearing belongs to 28.75HZ～29.95HZ.

In experiment, bearing have four running states, the condition of motor load power for 0 KW and speed for 1797 r/min is defined as S1, the condition of motor load power for 0.735 KW and speed for 1772 r/min is defined as S2, the condition of motor load power for 1.47 KW and speed for 1750 r/min is defined as S4, the condition of motor load power for 2.205 KW and speed for 1725 r/min is defined as S1. Bearing’s fault have four different degree, fault size for 0.1778mm is defined as f1, fault size for 0.3556 mm is defined as f2, fault size for 0.5334 mm is defined as f3, fault size for 0.7112 mm is defined as f4. Table 1 deacibes the the samples distribution in four runing states and four different fault degree.

Figure 1. Rolling bearing test experiment stand

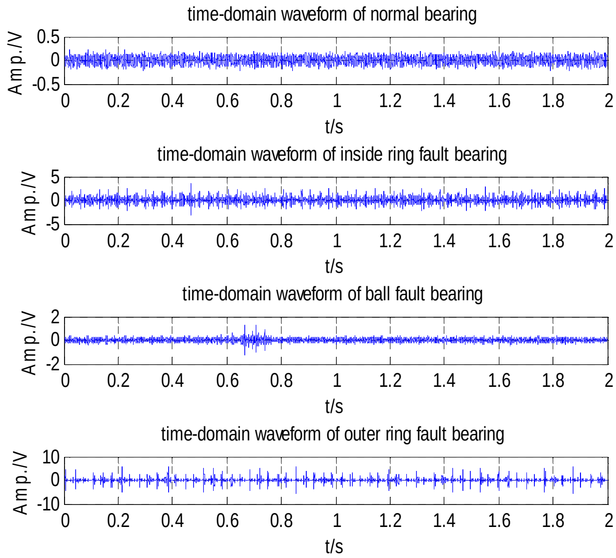

Figure 2. Time-domain waveform of four fault state signa

TABLE I.
THE SAMPLE DISTRIBUTION IN DIFFERENT STATES FOR ROLLING BEARING

No	Normal			Inner fault			Outer fault			Ball fault		
	State code	number		State code	number		State code	number		State code	number	
		Training sample	Test sample		Training sample	Test sample		Training sample	Test sample		Training sample	Test sample
1	S1	16	4	S1+f1	4	1	S1+f1	4	1	S1+f1	4	1
2	S2	16	4	S2+f1	4	1	S2+f1	4	1	S2+f1	4	1
3	S3	16	4	S3+f1	4	1	S3+f1	4	1	S3+f1	4	1
4	S4	16	4	S4+f1	4	1	S4+f1	4	1	S4+f1	4	1
5				S1+f2	4	1	S1+f2	4	1	S1+f2	4	1
6				S2+f2	4	1	S2+f2	4	1	S2+f2	4	1
7				S3+f2	4	1	S3+f2	4	1	S3+f2	4	1
8				S4+f2	4	1	S4+f2	4	1	S4+f2	4	1
9				S1+f3	4	1	S1+f3	4	1	S1+f3	4	1
10				S2+f3	4	1	S2+f3	4	1	S2+f3	4	1
11				S3+f3	4	1	S3+f3	4	1	S3+f3	4	1
12				S4+f3	4	1	S4+f3	4	1	S4+f3	4	1
13				S1+f4	4	1	S1+f4	4	1	S1+f4	4	1
14				S2+f4	4	1	S2+f4	4	1	S2+f4	4	1
15				S3+f4	4	1	S3+f4	4	1	S3+f4	4	1
16				S4+f4	4	1	S4+f4	4	1	S4+f4	4	1

In experiment, sampling frequency is 12kHz, bearings have four running states of normal state, inner ring fault, outer ring fault and ball fault. Figure 2 is the time-domain waveform of acceleration signal acquainted from drive-side bearing. When motor turns each cycle, sensor sampling points are 401～407. Let the dimension of sample space be 417, if the amount is less than 417, then padded with 0.For each fault type, the amount of training sample vector are 64 and test sample vector are 16. So vector matrix corresponding to them are respectively 256×417 and 64×417 .

Estimation of intrinsic dimension are 19.5849 by maximum likelihood estimation. In the experiment, the intrinsic dimension adopts 12, 15, 18 and 20. Parameter σ of Gaussian kernel function was taken as 0.01, penalty factor C as 0.1, supervised operator as 0.3.

Figure 3 describes four methods' fault diagnosis result to training and test samples in different intrinsic dimension condition. In figure1, left column chart describes the result of fault diagnosis four methods to training samples; right column chart describes the result of fault diagnosis four methods to test samples. Solid line represent the two supervised methods, dotted line represent two non-supervised method. '*' means the diagnosis' result of methods based on LLTSA-SVM, '○' means the result of methods based on ILTSA- SVM,

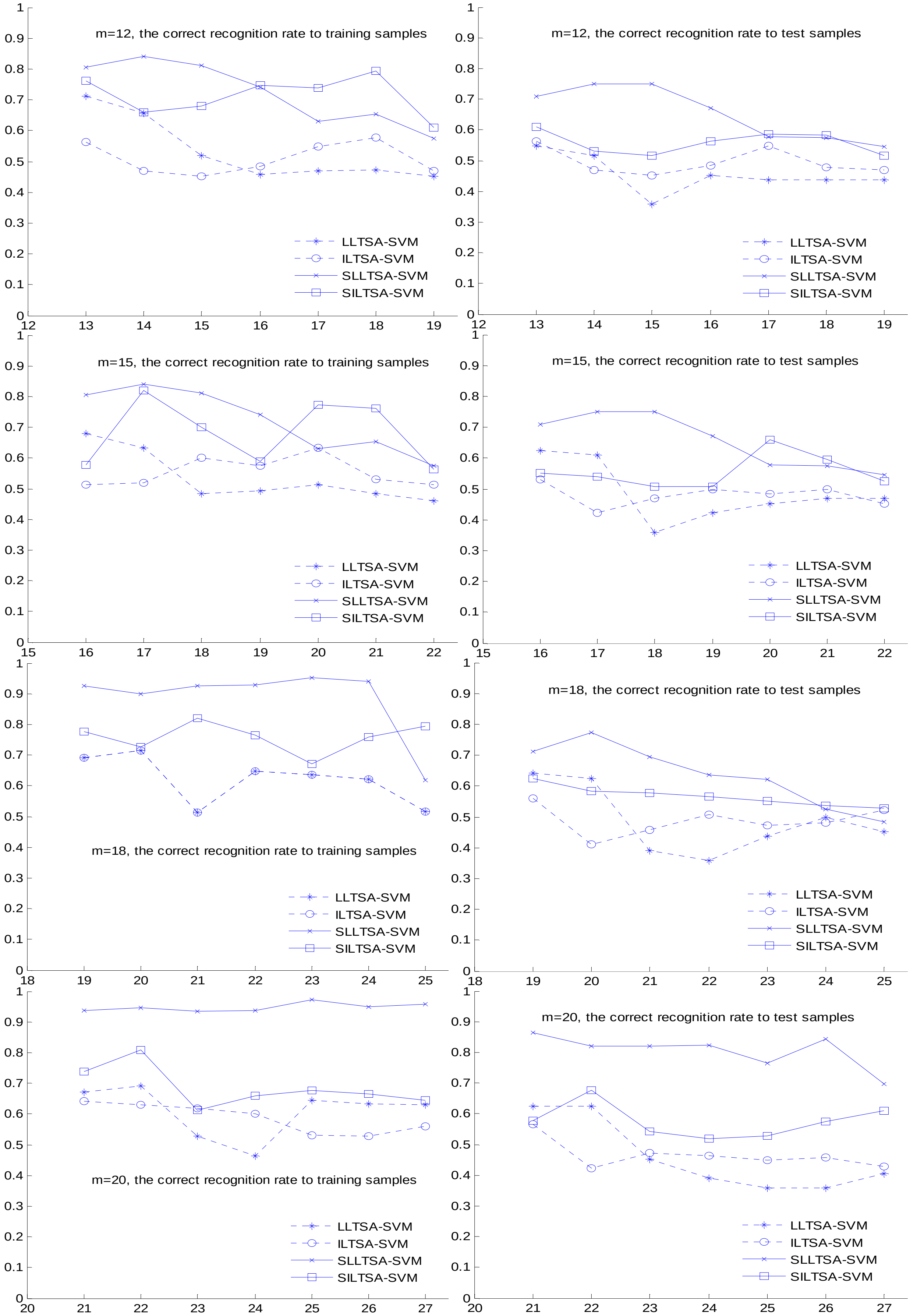

Figure 3. Four methods' fault diagnosis result to training (left) and test (right) samples

TABLE II.
THE HIGHEST CORRECT RECOGNITION RATE IN DIFFERENT INTRINSIC DIMENSION CONDITION

	m=12		m=15		m=18		m=20	
	Training sample	Test sample	Training sample	Test sample	Training sample	Test sample	Training sample	Test samples
ILTSA-SVM	64.45%	57.81%	63.28%	53.13%	62.89%	55.86%	64.04%	56.64%
SILLTSA-SVM	79.30%	60.94%	77.34%	66.02%	82.03%	62.50%	80.86%	67.58%
LLTSA-SVM	71.09%	54.69%	67.97%	62.50%	71.48%	64.06%	69.14%	62.50%
SLLTSA-SVM	83.98%	75.00%	84.38%	79.30%	92.58%	77.34%	95.70%	86.33%

' x ' means the result of methods based on SLLTSA-SVM, '☐'means the result of methods based on SILTSA-SVM. Table 1 describes the highest correct recognition rate obtained by four multi-layer kernel learning methods in different intrinsic dimension condition. Either figure 3 or Table 1, either training sample or test sample, the results show that fault diagnosis methods based on supervised learning have better effect than non-supervised learning.

VI. CONCLUSION

To make full of the advantage of manifold learning and support vector machine respectively, we propose two multi-layer kernel learning model, ILTSA-SVM and ILTSA-SVM. The method resolves the generalization problem of nonlinear manifold learning, but also enhances its ability of fault diagnosis. At the same time, the supervised idea is introduced to two multi-layer kernel learning algorithms; SILTSA-SVM and SILTSA-SVM are proposed and may solve the problem not to effectively extract the fault characteristics of mechanical equipment in two non-supervised methods.

The experiment result shows that four methods can obtain result of the bearing fault diagnosis in some extent, but two supervised multi-layer kernel learning have better effect than others. In four methods, the multi-layer kernel learning methods based on LLTSA have better effect than method based on ILTSA, it because that LLTSA could bring the certain projection error by means of replacing non-linear mapping with linear mapping, but it does not change the local structure of the data space. While in ILTSA, test samples and training samples are mixed, then recalculate neighbor matrix. In the new neighborhood, some point’s relative position changes, and error will be Is magnified in the projection process. Eventually lead to reduce the effect of fault diagnosis.

ACKNOWLEDGMENT

Financial support from National Natural Science Foundation of China (51175170), The Industrial Cultivation Program of Scientific and Technological Achievements in Higher Educational Institutions of Hunan Province (10CY008), Natural Science Foundation of Hunan Province Key Project (09JJ8005), Hunan Province Science Plan Project (2010gk3042), Aid program for science and technology innovative research team in higher educational institutions of Hunan province, are gratefully acknowledged.

REFERENCES

[1] H.S. Seung, D.L. Daniel. "The manifold ways of perception", *Science*.vol.290, pp.2268-2269,2000.
[2] S. Roweis, L. Saul. "Nonlinear dimensionality reduction by locally linear embedding", *Science*.vol.290, pp.2323-2326, 2000.
[3] M. Belkin, P. Niyogi. "Laplacian eigenmaps for dimensionality reduction and data representation", *Neural Computation*. vol.15, pp.1373-1396,2003.
[4] Z. Y. Zhang, H. Y. Zha. "Principal manioflds and nonlinear dimension reduction via Tangent Space Alignmnet", *SIAM Journal on Scientific Computing*. vol.26, pp.313-338, 2005.
[5] X. He, P. Niyogi. "Locality preserving projections", *Proceedings of Advances in Neural Information Processing System*, Vancouver: Canada, 2003.
[6] X. He, D. Cai, S. Yan and H.J. Zhang. "Face recognition using Laplacianfaces", *Proceedings of the 10 IEEE International Conference on Computer Vision*, Beijing: China, 2005. pp: 1208–1213.
[7] T. H. Zhang, Y. Jie and D.L. Zhao. "Linear local tangent space alignment and application to face recognition", *Neuro computing Letters*, vol.70, pp.1547- 1553, 2007.
[8] M. H. Law, A. K. Jain. "Building k-conneted neighborhood graphs for isometric data embedding", *IEEE Trans, Pattern Analysis and Machine Intelligence*, vol.28, pp.377-391, 2006.
[9] O. Kouropteva, O. Okun and N. IPietik. "Matti Pietikäinen. Supervised locally linear embedding algorithm for pattern recognition", *Pattern recognition*. Vol.38, pp.1764-1767, 2005.
[10] X. M. Liu, J. W. Yin and Z. L. Feng. "Ubcrenebtak incremental manifold learning via tangent space alignment ",*Artificial Neural Networks in Pattern Recognition, Ulm*: Germany, 2006,pp.107-121
[11] H. Choi, S.Choi. "Kernel isomap", *Electronics Letters*, vol.40, pp.1612-1613, 2004.
[12] G.B. Wang, Y. L. Liu and L. P. Huang. "Fault diagnosis based on kernel schur-orthogonal local fisher discriminate ", *Chinese Journal of Scientific Instrument*. vol.31, pp. 1005-1009
[13] K.Q. Weinberger, F. Sha and L.K. Saul. "Learning a kernel matrix for nonlinear dimensionlity reduction" *Proc of the 21st International Conference on Machine Learning*. pp.839-846, 2004
[14] S. C. Yan, D. Xu and H. J. Zhang. Graph embedding a general framework for dimensionality reduction", *Proc of Conference on Computer Vision and Pattern Recognition*. pp. 830-837, 2005.
[15] Qihong Huang, "Research on manifold learning: theories and applications in images", *University of Electronic Scinece and technology of china*, Doctoral Dissertations, 2007

[16] L. Saul, S. Roweis. Think globally, fit locally: unsupervised learning of low dimensional manifolds", *Journal of Machine Learning Research*, vol.4, pp. 119-155, 2002.
[17] Information on http://www.eecs.case.edu/laboratory/bearing.

Guangbin Wang was born in December 1974 in Lanzhou city of Henan province. He obtained PHD in 2010 from Central South University, research field are manifold learning and equitment's fault diagnosis.

He has been working at Hunan University of Science and Technology since 1999, mainly engaged in student management, teaching and scientific research. His research interest are nonlinear signal processing and equitment's condition monitoring and fault diagnosis.

Yilin He was born in 1988; she is graduate candidate in Key Laboratory of Health Maintenance for Mechanical Equipment of Hunan Province. Her mainly research interests include manifold learning and equipment's fault diagnosis.

Kuanfang He, male, born in 1979, received master degree in 2006 and PHD in 2009 from South China University of Technology, research field are dynamic monitoring and control for electromechanical systems. He has been working at Hunan University of Science and Technology since 2009, mainly engaged in theory and practice of teaching, scientific research. His research interest is dynamic monitoring and control for electromechanical systems.

Dr. He is a member of Hunan Province instrumentation Institute. In recent years, he has presided one National Natural Science Foundation of China and two provincial research projects, obtained three provincial academic reward, published more than 20 academic papers.