Engineering Applications of Artificial Intelligence 50 (2016) 60–70

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Incremental supervised locally linear embedding for machinery fault diagnosis

Yuanhong Liu [a,b], Yansheng Zhang [a,b,*], Zhiwei Yu [b], Ming Zeng [b]

[a] School of Information and Electrical Engineering, Northeast Petroleum University, Daqing 163318, PR China
[b] Space Control and Inertial Technology Research Center, Harbin Institute of Technology, Harbin 150001, PR China

ARTICLE INFO

Article history:
Received 12 June 2015
Received in revised form
5 November 2015
Accepted 15 December 2015
Available online 21 January 2016

Keywords:
Block matrix decomposition
Dimension reduction
Linear discriminant analysis
Locally linear embedding
Fault diagnosis

ABSTRACT

Locally linear embedding (LLE) is a promising algorithm for machinery fault diagnosis, but LLE operates in a batch mode and lacks discriminant information, which lead to be negative for fault diagnosis. In this paper, incremental supervised LLE (I-SLLE) is investigated for submersible plunger pump fault diagnosis. In the I-SLLE algorithm, block matrix decomposition technology is introduced to deal with out-of-sample data, while a part of old low-dimensional coordinates is also updated, upon which an iterative method is presented to update all the data for refining the accuracy. Furthermore, in order to improve the classification capability of LLE, discriminant information is assembled in the cost function of LLE. Based on I-SLLE, a new machinery fault diagnosis method is proposed. At first, I-SLLE is utilized to extract the feature of an original dataset, and then support vector machine (SVM) is employed to classify the test data in the feature space. Experiments on synthetic datasets as well as real world datasets are performed, demonstrating the efficiency and the accuracy of the proposed algorithm.

1. Introduction

Submersible plunger pump lifting system is a new oil recovery equipment used in the oil field. Its signal dimensionality is very high and the signal has strong non-stationary characteristics. Usually, residual-based method (Serdio et al., 2014a; Yin and Huang, 2015), gradient information method (Serdio et al., 2015), time-domain index method (Yin et al., 2015a; Tabatabaeipour, 2015), frequency-domain index method (Zhang et al., 2015), or time/frequency-domains index method (Wang et al., 2015) etc. is employed to recognize the fault type of an equipment in the original space. However, as we all know, this kind of method will fail in a high-dimensional space. Hence, conventional fault diagnosis methods are usually inefficient for submersible plunger pump. A widely used way to resolve this problem is dimension reduction methods (Ali et al., 2015; Yin et al., 2014), i.e., project an original dataset into a low-dimensional feature space, and the running state of an equipment can be diagnosed in the low-dimensional space (Jin et al., 2014; Yu, 2013; Serdio et al., 2014b).

Recently, dimension reduction algorithms for machinery fault diagnosis have been intensively investigated (Ferracuti and Giantomassi, 2015; Echevarra et al., 2014; Yin et al., 2015b). Principal components analysis (PCA) (Zhou et al., 2014) and multi-dimensional scaling (MDS) (Lee and McDonnell, 2014) are two classical dimension reduction algorithms, and they are mutually dual form, which makes them to have similar characteristics. For their linear mode property, when the distribution of a dataset is nonlinear, PCA and MDS may be invalid. In addition, they all belong to unsupervised learning methods, in other words, class information cannot be utilized for dimension reduction, weakening the recognition accuracy. Therefore, some nonlinear dimension reduction algorithms are proposed. In the nonlinear algorithms, the local high-dimensional space is assumed to be linear, and dimensionality is reduced by preserving the local property of a high-dimensional dataset in low one. Nonlinear dimension reduction methods can effectively show the real structure of the original dataset, which is advantage for fault diagnosis. Currently, representative nonlinear algorithms mainly include isometric mapping (ISOMAP) (Tenenbaum et al., 2000), local tangent space alignment (LTSA) (Zhang and Zha, 2002), diffusion map (Fernández et al., 2015), kernel principal component analysis (KPCA) (Shao et al., 2014), and locally linear embedding (LLE) (Roweis and Saul, 2000), etc. Among these algorithms, LLE is an effective one and has been widely used. In the LLE algorithm, the local geometry of original dataset is preserved in embedding space, which provides a better understanding of the internal structure of the high-dimensional space. Moreover, LLE can avoid local minima and only few free parameters need to be set. Although LLE has so many

* Corresponding author at: School of Information and Electrical Engineering, Northeast Petroleum University, Daqing 163318, PR China.
E-mail addresses: liuyuanhong@nepu.edu.cn (Y. Liu), zhangyansheng@nepu.edu.cn (Y. Zhang), yzw@hit.edu.cn (Z. Yu), zenmging@hit.edu.cn (M. Zeng).

http://dx.doi.org/10.1016/j.engappai.2015.12.010
0952-1976/

advantages, like most of the other nonlinear algorithms, LLE has two obvious drawbacks for feature extraction: (1) LLE does not provide an explicit mapping from a high-dimensional space to a feature space, i.e., for newly coming samples, the old data augmented by the new samples must be completely recalculated by LLE, which is computationally intensive. (2) LLE performs in an unsupervised manner, namely, class information cannot be used in the LLE.

In order to solve these two problems, some algorithms based on LLE have been investigated. Yoshua et al. (2003) provided a unified framework for extending LLE, Isomap, LE, and MDS, in which a new sample can be calculated by a kernel function. However, the accuracy of this method is mainly dependent on the estimation of a dataset distribution. Elgammal and Lee (2007) employed RBF to learn a mapping from an embedding space to an input space, and an approximate solution for the inverse mapping can be obtained in a closed form, by which new samples can be easily computed. However, the computational cost is expensive. Kouropteva et al. (2005) proposed a non-parametric incremental LLE algorithm that is referred to as ILLE1. In the ILLE1, an explicit mapping was constructed by the high-dimensional neighbors and the low-dimensional neighbors of a new sample. A similar algorithm that is referred to as ILLE2 was raised in Lawrence and Sam (2003), in which the low-dimension coordinate of a new data can be acquired by a linear combination of its low-dimensional neighbors. In actually, ILLE1 and ILLE2 are equivalent in essence, and they can only deal with new data points just one by one. Su et al. (2014) adopted block matrix decomposition technology to simplify the calculation of the cost matrix M to obtain the low-dimensional coordinates of the new samples, which is referred to as ILLE3. ILLE3 is the batch mode of ILLE1 and ILLE2. As mentioned above, the existing incremental LLE methods only concern about the embedding results of new samples, however, the meaningful information of the new samples is not used to improve the accuracy of old data, and their results are rough for they all belong to local optimal algorithms. In addition, all these incremental LLE algorithms belong to unsupervised learning methods, that is, the class information of dataset is not used to aid in dimension reduction. For solving this issue, discriminant information has been considered in LLE. A supervised LLE was proposed in the literature (de Ridder et al., 2003), in which the class information was used to estimate the similarity of pairwise. To maximize the separability of data among different classes, Zhang et al. (2004) suggested that LDA should be employed to treat the result obtained by LLE, upon which the geometry information of original dataset is destroyed, and its computational complexity is also very high.

In this paper, the I-SLLE algorithm is proposed by improving literature (Su et al., 2014; Zhang et al., 2004). Based on the I-SLLE algorithm, a novel fault diagnosis method for submersible plunger pump is presented. The main contributions of this paper are summarized as follows:

1. Submersible plunger pump lifting system is a new oil recovery equipment used in the oil field. To the best of my knowledge, the nonlinear dimension reduction method is firstly employed to deal with the signal of submersible plunger pump for fault diagnosis and it has been demonstrated to be effective by our experiments.
2. To use the meaningful information of the new samples, in our proposed method, the new samples and the old data are simultaneously updated, while the algorithm introduced in the literature (Su et al., 2014) only concerned about the new samples. At first, with the new samples coming, the data whose neighbors are changed along with the new samples will be calculated, and then an iterative method is introduced to update all the data. Compared with the literature (Su et al., 2014), the accuracy of our proposed method is more higher, which is demonstrated by our experiment.
3. LDA is introduced in LLE to improve the classification capability. The algorithm proposed in the literature (Zhang et al., 2004) dealt with dataset by LLE and LDA, respectively, while we add LDA as an item in the cost function of LLE, remaining the structure information and discrimination information and reducing the computational complexity.

The remainder of this paper is organized as follows: Section 2 briefly reviews LLE, and our I-SLLE algorithm is introduced in detail. In Section 3, based on the I-SLLE algorithm, a novel fault diagnosis method for submersible plunger pump is described. In Section 4, experiments on synthetic and real world datasets demonstrate that I-SLLE is effective. Finally, some conclusions and further research directions are drawn in Section 5.

2. Incremental supervised locally linear embedding

2.1. Locally linear embedding

Let $X=[x_1,x_2,\ldots,x_N]\in R^{D\times N}$ represent a high-dimensional dataset, where $x_i\in R^D$ denotes the ith sample, for $i=1,2,\ldots,N$, and N represents the sample size. LLE aims to map X into $Y=[y_1, y_2,\ldots,y_N]\in R^{d\times N}$ ($d\ll D$), where Y is the low-dimensional coordinate of X. The implementation procedure of LLE can be conducted with the following three steps (Su et al., 2014):

1. Find the neighbors for each point by using the K-NN method. For a point x_i, let $X^i=[x_i^1,x_i^2,\ldots,x_i^k]\in R^{D\times k}$ denote the neighbors of x_i, where k is the number of neighbors. First, a similarity measure of pairwise in dataset is calculated, then k points that are most similar to x_i are chosen as the neighbors of x_i. Although Euclidean distance is the most common measure criteria, it fails in high-dimensional space. Lately, some new similarity measure methods have been introduced such as Minkowski distance, correlation, hamming distance, and spearman's rank correlation etc., which should be selected according to the distribution of a dataset. Moreover, the parameter k plays an important factor for the final result in LLE. If k is too large, then the points with different types may be involved into the neighbor, for which the local information may be lose; if k is too small, then the points with the same class cannot be sufficient overlap, therefore, the global structure will be destroyed. The result is stable over a wide range of k, but does break down as k becomes too small or too large. The selection of k can be referred to the literature (Hu et al., 2014).

2. Calculate the reconstruction weights. Let $W_i=(w_{i1},w_{i2},\ldots,w_{ik})'\in R^{k\times 1}$ represent the reconstruction weight of x_i. It can be determined by solving the constrained least squares (LS) problem:

$$\arg\min\left\|x_i-\sum_{j=1}^{k}w_{ij}x_i^j\right\|_2^2=\arg\min\left\|x_i-X^iW_i\right\|_2^2$$
$$\text{s.t.}\quad e_k'W_i=1 \tag{1}$$

where $e_k=[1,1,\ldots,1]'\in R^{k\times 1}$. The constrained condition can ensure that the solution of Eq. (1) is translation invariance. According to the Lagrange multiplier, when the partial derivative of Eq. (1) with respect to W_i is zero, the solution is minimum. Employing the Lagrange multiplier on Eq. (1) and setting the result to zero lead to

$$X^i(x_i-X^iW_i)+\beta e=0$$

where β is a Lagrange multiplier that can modify the quadratic form to satisfy the constraint condition. Then the optimal solution W_i can be

computed as follows:

$$W_i = \frac{G_i^{-1}e}{e'G_i^{-1}e}$$

where $G_i = (x_i - x_i^1, x_i - x_i^2, \ldots, x_i - x_i^k)'(x_i - x_i^1, x_i - x_i^2, \ldots, x_i - x_i^k) \in R^{k\times k}$. If k is smaller than D, then G_i^{-1} is existent and unique, W_i being stable; if k is larger than D, then G_i is singular or nearly singular, which makes W_i sensitive to noise. Usually, in this case, a regularization item should be added in G_i (Wang, 2014),

$$G_i = G_i + \frac{\Delta^2}{k}\mathrm{tr}(G_i)I_k$$

where $\mathrm{tr}(G_i)$ represents the trace of G_i, $I_k \in R^{k\times k}$ is an identity matrix, Δ is a constant and $\Delta \ll 1$. Note that W_i is sensitive to Δ.

3. Compute the low-dimensional coordinate Y based on the reconstruction weights by minimizing the cost function:

$$\begin{aligned} &\arg\min \sum_{i=1}^{N}\left\| y_i - \sum_{j=1}^{k} w_{ij}y_i^j \right\|_2^2 = \mathrm{tr}(Y'MY) \\ &\text{s.t.} \quad \frac{1}{N}Y'Y = I_N \\ &\quad Y'e_N = 0 \end{aligned} \tag{2}$$

where $M = (I_N - W)'(I_N - W)$, and $W \in R^{N\times N}$ represents weight coefficient matrix. The ith row and the jth column element of the matrix W denotes the weight value between x_i and x_j. If x_i is not a neighbor of x_j, then $w_{ij} = 0$. The aforementioned two constraints make the solution of the cost function be invariant to translations and rescalings. The low-dimensional coordinate of X can be obtained by calculating the bottom $d+1$ eigenvectors of the matrix M. Since the normalization constraint $\sum_{j=1}^{k} w_{ij} = 1$, zero is a trivial eigenvalue that should be excluded, so the remaining d eigenvectors are the final low-dimensional coordinates. These d smallest eigenvalues are usually very small, but the largest eigenvalue is very large, which will lead to ill-posed problem, i.e., it is too difficult to obtain the accurate small eigenvalues and their corresponding eigenvectors. Furthermore, M is a large sparse symmetric matrix, and its computation costs most of time in LLE. In addition, the selection of d is a key for fault diagnosis. The dimension d represents the number of features to characterize a signal. A large d means more information, whereas a small d can simplify the complexity of a dataset. Usually, d can be estimated by using geodesic minimal spanning tree (GMST), maximum likelihood estimate (MLE), and fractal method, among others (Liu et al., 2015).

2.2. Incremental locally linear embedding

From the analysis of LLE, one can see that LLE does not provide an explicit mapping from an original space to a low-dimensional feature space. Completely recomputing LLE for out-of-sample data is very expensive. The computational cost of every step of LLE is listed in Table 1. From Table 1, we can see that the decomposition of the cost matrix M spends the most of time. Therefore, it is necessary to simplify M for improving the capability of LLE to deal with new samples.

Table 1
Each step computational cost of LLE.

Step	Action performed	% of runtime
1	Find k neighbors	24
2	Calculate reconstruction weights matrix	17
3	Calculate low-dimensional coordinate	59

Su et al. (2014) suggested that the matrix M should be simplified by using block matrix decomposition technology, in which only the elements relating with the new samples are employed to calculate the low-dimensional coordinates of the new samples. Although the calculational speed is improved in this method, the useful information of the new samples is not fully utilized, and the result is also rough. Hence, in this paper, a new incremental LLE (I-LLE) based on block matrix decomposition technology is proposed. In the I-LLE algorithm, the new samples are effectively treated, at the same time the meaningful information of the new samples is also considered to update the old data whose neighbors are changed by the new samples. Furthermore, for obtaining a more accurate result, an iterative method is introduced to refine the rough solution. Let $\tilde{Y} = [Y_{unchange}, Y_{change}, Y_{new}]$ be a low-dimensional coordinate, where $Y_{unchange} \in R^{d\times N_1}$ denotes the original low-dimensional points whose neighbors are not changed, $Y_{change} \in R^{d\times N_2}$ represents the points that their neighbors contain the new samples, $Y_{new} \in R^{d\times N_3}$ indicates the low-dimensional coordinate of the new samples, where N_1, N_2, and N_3 indicate the sample number of $Y_{unchange}$, Y_{change}, and Y_{new}, respectively, d represents the dimensionality of the low-dimensional coordinate. Only Y_{change} and Y_{new} should be calculated. For convenience, we set $Y_{update} = [Y_{change}, Y_{new}]$. Rewrite Eq. (2) by using the block matrix decomposition technology and it can be expressed as follows:

$$\arg\min \mathrm{tr}\left([Y_{unchange}, Y_{update}] \begin{bmatrix} E & F \\ G & H \end{bmatrix} \begin{bmatrix} Y'_{unchange} \\ Y'_{update} \end{bmatrix} \right) \tag{3}$$

where $M = \begin{bmatrix} E & F \\ G & H \end{bmatrix}$, $E \in R^{N_1\times N_1}$, $F \in R^{N_1\times(N_2+N_3)}$, $G \in R^{(N_2+N_3)\times N_1}$, $H \in R^{(N_2+N_3)\times(N_2+N_3)}$. Eq. (3) is a quadratic form, therefore, E, F, G, and H can be easily determined by the property of the quadratic form. Now, Eq. (3) can be unfolded as follows:

$$\begin{aligned} &\arg\min \mathrm{tr}(Y_{uchange}EY'_{change}) + \mathrm{tr}(Y_{uchange}FY'_{update}) + \mathrm{tr}(Y_{update}GY'_{unchange}) \\ &\quad + \mathrm{tr}(Y_{update}HY'_{update}) \end{aligned} \tag{4}$$

we calculate the partial derivative of Eq. (4) with respect to Y_{update} and set the result to zero:

$$\mathrm{tr}(Y_{unchange}F) + \mathrm{tr}(Y_{unchange}G') + \mathrm{tr}(Y_{update}H + Y_{update}H') = 0 \tag{5}$$

The above equation is unconstrained, and only Y_{update} is unknown. Hence, Y_{update} can be directly solved as follows:

$$Y_{update} = -(Y'_{unchange}F + Y_{unchange}G')(H + H')^{-1} \tag{6}$$

computing the inverse of $(H+H')$ is time-consuming. Furthermore, if it is singular, then the result is sensitive to noise. Hence, SVD is introduced to calculate the matrix inverse, and the formula can be described as follows:

$$H + H' = [U_1, U_2 \ldots U_{(N_2+N_3)}] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{(N_2+N_3)} \end{bmatrix} \begin{bmatrix} V'_1 \\ V'_2 \\ \vdots \\ V'_{(N_2+N_3)} \end{bmatrix}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{(N_2+N_3)}$ are the singular eigenvalues of $H+H'$, $[U_1, U_2 \ldots U_{(N_2+N_3)}]$ and $[V_1, V_2 \ldots V_{(N_2+N_3)}]'$ are column and row orthogonal basis, respectively. Only the largest n eigenvalues and their corresponding eigenvectors are employed to compute the inverse of $H+H'$, and then the inverse can be calculated as follows:

$$[H + H']^{-1} = [V_1, V_2 \cdots V_n] \begin{bmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda_n} \end{bmatrix} \begin{bmatrix} U'_1 \\ U'_2 \\ \vdots \\ U'_n \end{bmatrix}.$$

Till now, the low-dimensional coordinate of the original dataset is obtained, but the result is rough for the reason of local optimum. The accuracy may be high when the size of new samples is very small and far less than that of the old ones, but when this condition is not satisfied, the result is rough. Hence, it is necessary to refine the result by using the known information. Since the rough result is known, we can use an iterative method to update it. Gradient descent is the most simple iterative method, and it can be shown as follows:

$$Y(i+1) = Y(i) - \gamma \nabla_Y \phi(Y(i)) \tag{7}$$

where i represents the ith iterative number, γ denotes a small positive number known as the learning rate, ∇_Y indicates the gradient operator with respect to Y, $\phi(Y)$ denotes the function with respect to Y. In this paper, $\phi(Y)$ represents Eq. (3). It is well known that such a learning scheme can be very slow, but in our method, the initial estimate $Y(0)$ is known, and it is close to the optimal solution, by which this algorithm can rapidly converge. Thus, the complete algorithm can be described as follows:

Algorithm 1. Refining the low-dimensional coordinates.

1. Calculate new samples and update part of old data by using block matrix decomposition technology, and then obtain the rough result.
2. Set the rough result as the initial value of the iterative method.
3. Implement Eq. (7), until the stop condition is satisfied.

For assessing the quality of the update, the residual variance is introduced (Kouropteva and Okun, 2002). It is defined as $f = 1 - \rho_{D_x D_y}$, where ρ is the correlation coefficient, D_x and D_y are similarity measures of high-dimensional data and low-dimensional data, respectively. A small f means an excellent dimension reduction result. If the value of f is good, then the update should be accepted, otherwise, this update should be discarded.

2.3. Supervised locally linear embedding

As shown in Eq. (6), Y_{update} has a great relationship with $Y_{unchange}$. If $Y_{unchange}$ contains superior discrimination information, Y_{update} may be advantage for fault diagnosis. Unfortunately, $Y_{unchange}$ only holds structure information of original dataset in the LLE algorithm. Local discriminant analysis (LDA) (Wang and Sun, 2013) is a supervised learning algorithm, which can maximize between-class scatter and minimize within-class scatter, that is, discriminant information can be provided by LDA. Zhang et al. (2004) suggested that LDA should be used to reduce the dimensionality of the data that had been treated by LLE. Although this method can improve the discrimination of different classes, it will introduce two problems: (1) the structure information will be destroyed by LDA; (2) since both LLE and LDA are respectively implemented, the computational cost is expensive. In order to resolve the two problems, we integrate LDA into LLE. Let S^w and S^b be the within-class and between-class scatter matrix, respectively. S^w and S^b can be defined as follows:

$$S^w = \frac{1}{n_i} \sum_{i=1}^{m} \sum_{j=1}^{n_i} (x_j - \mu_i)'(x_j - \mu_i)$$

$$S^b = \sum_{i=1}^{m} (\mu_i - \mu)'(\mu_i - \mu)$$

where m denotes the number of categories, n_i represents the sample size of the ith category, μ_i indicates the mean of samples in the ith category, and μ denotes the mean of all samples. We add S^w and S^b in Eq. (2), and minimize the following objective function:

$$\alpha \sum_{i=1}^{N} \left\| y_i - \sum_{j=1}^{k} w_{ij} y_i^j \right\|_2^2 + (1-\alpha) \left(\frac{1}{n_i} \sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_j - \mu_i)'(y_j - \mu_i) - \sum_{i=1}^{m} (\mu_i - \mu)'(\mu_i - \mu) \right) = \alpha Y' M Y + (1-\alpha)(Y' L_w Y + Y' L_b Y)$$
$$= Y'(\alpha M + (1-\alpha)(L_w + L_b))Y \tag{8}$$

where α is used to balance the trade-off between structure information and discriminant information, $L_w \in R^{(N \times N)}$ and $L_b \in R^{(N \times N)}$ denote the coefficient matrix of S^w and S^b, respectively. If α is equal to zero, then the method becomes LDA algorithm; if α is equal to one, then the method becomes the LLE algorithm. Therefore, the LLE and the LDA are the special case of our proposed method. The minimum d eigenvectors of the matrix $\alpha M + (1-\alpha)(L_w + L_b)$ is the optimal solution of Eq. (8). The solution of Eq. (8) implies both the local geometric structure information and the discriminant information, namely, the internal structure of original data is preserved, while the different classes can be effectively separated.

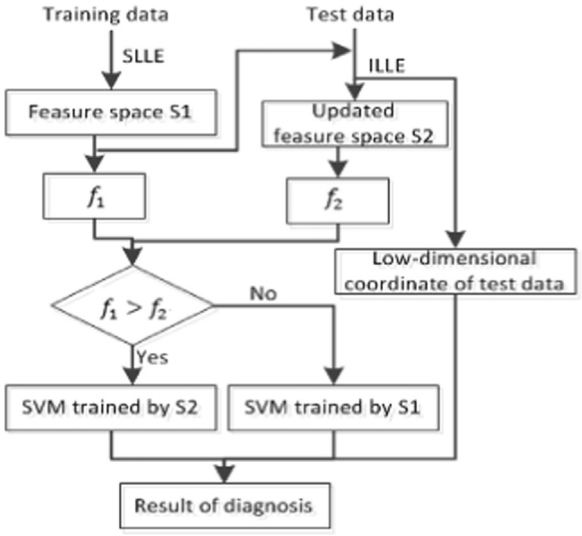


Fig. 1. Calculation chart of I-SLLE.

Table 2
Datasets operated on experiments.

Dataset	Sample size	Dimensionality	Category
S-curve	1500	3	
Punctured sphere	1500	3	
Toroidal helix	1500	3	
Iris	150	4	3
Motor bearing	400	1000	4
Submersible plunger pump	160	1800	4

3. Fault diagnosis by I-SLLE

Feature extraction is the core of fault diagnosis, further, the speed and the accuracy of feature extraction are important for fault identification. However, conventional LLE operates in a batch and unsupervised mode, which reduces the efficiency of LLE to extract feature. In order to overcome the intrinsic defects of LLE, we integrate I-LLE with SLLE (I-SLLE) to improve the performance of LLE. Based on the I-SLLE algorithm, a new fault diagnosis

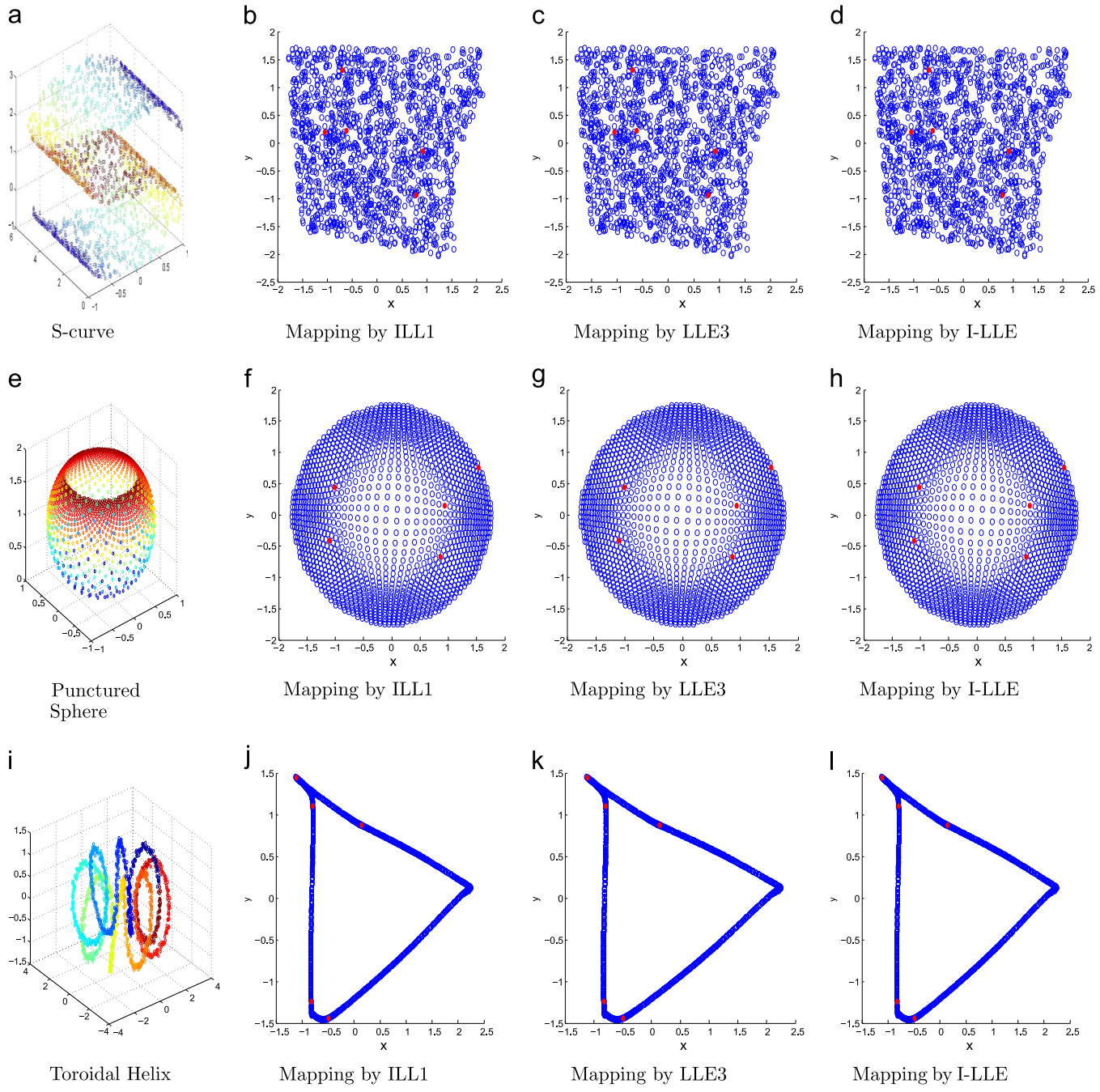


Fig. 2. Qualitative analysis of the incremental LLE. The blue points represent training data, and the red points denote test data. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

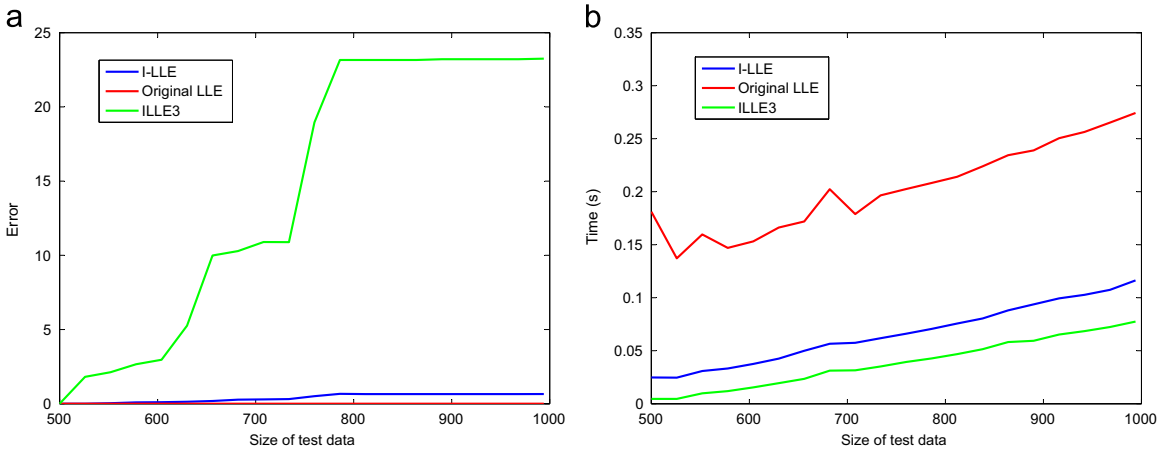


Fig. 3. Comparison of the three algorithms with different size of test data. (a) Error, (b) computing time.

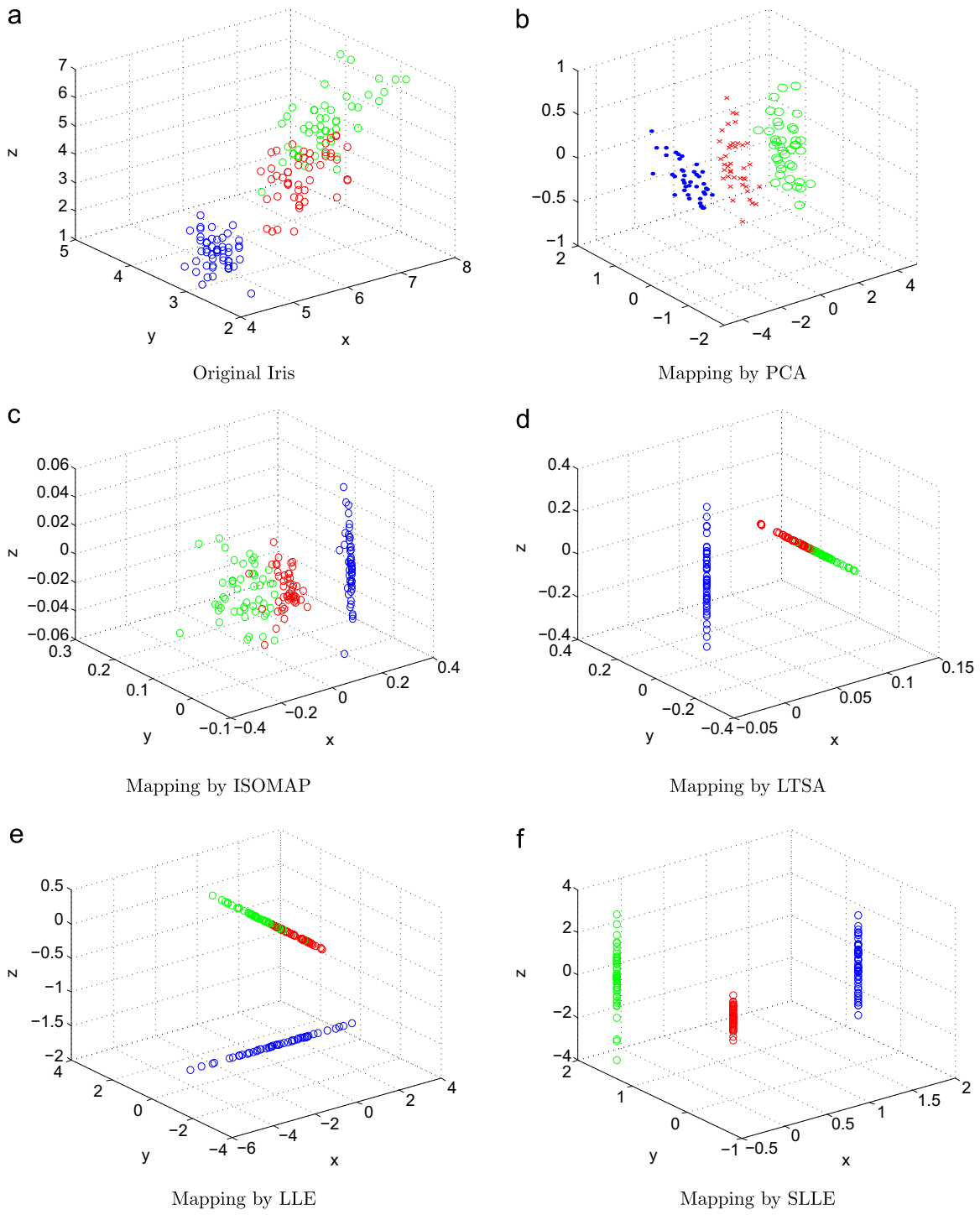


Fig. 4. Qualitative analysis of the incremental LLE. The blue points indicate Setosas data. The red points denote Virginicas data. The green points represent Versicolors data. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

method is proposed. At the beginning, training dataset with its class information is projected by SLLE for obtaining a low-dimensional feature space that has an excellent discrimination; then, calculate the new samples and update the whole dataset by using Algorithm 1; next the residual variance f is estimated for determining whether this update should be accepted; further, SVM is trained by the low-dimension coordinate of training data with their class information; at the end, the test samples are diagnosed by the trained SVM. The fault diagnosis chart based on I-SLLE is shown in Fig. 1, and the detailed description can be summarized as follows:

1. Set up training samples and test samples by pre-processing the observed signals sampled from a machinery.

2. Choose suitable parameters such as d, k, and α etc. In addition, similarity measure method should also be selected according to the distribution of a dataset.

3. SLLE is employed on the training dataset to obtain the low-dimensional coordinate, and the residual variance f_1 is calculated. Moreover, I-LLE is utilized to deal with the new samples, and then all the data is also updated.

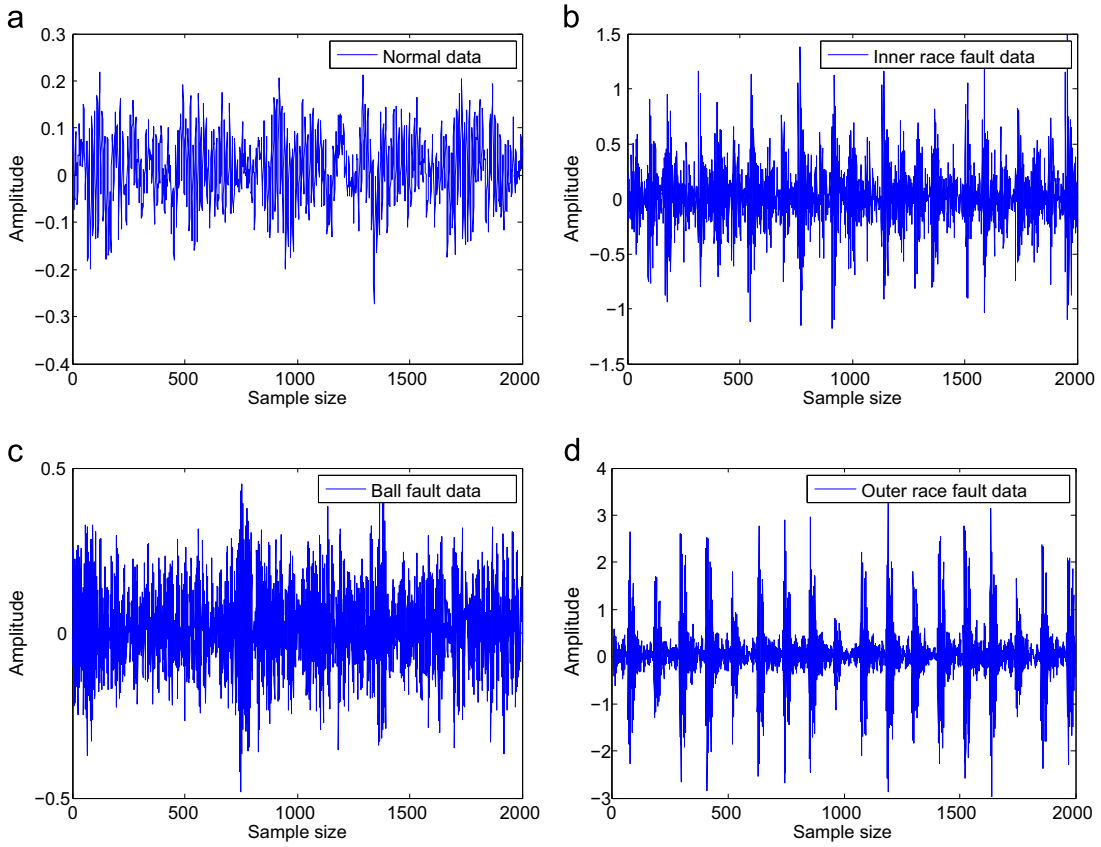


Fig. 5. Four typical vibration signals of motor bearing.

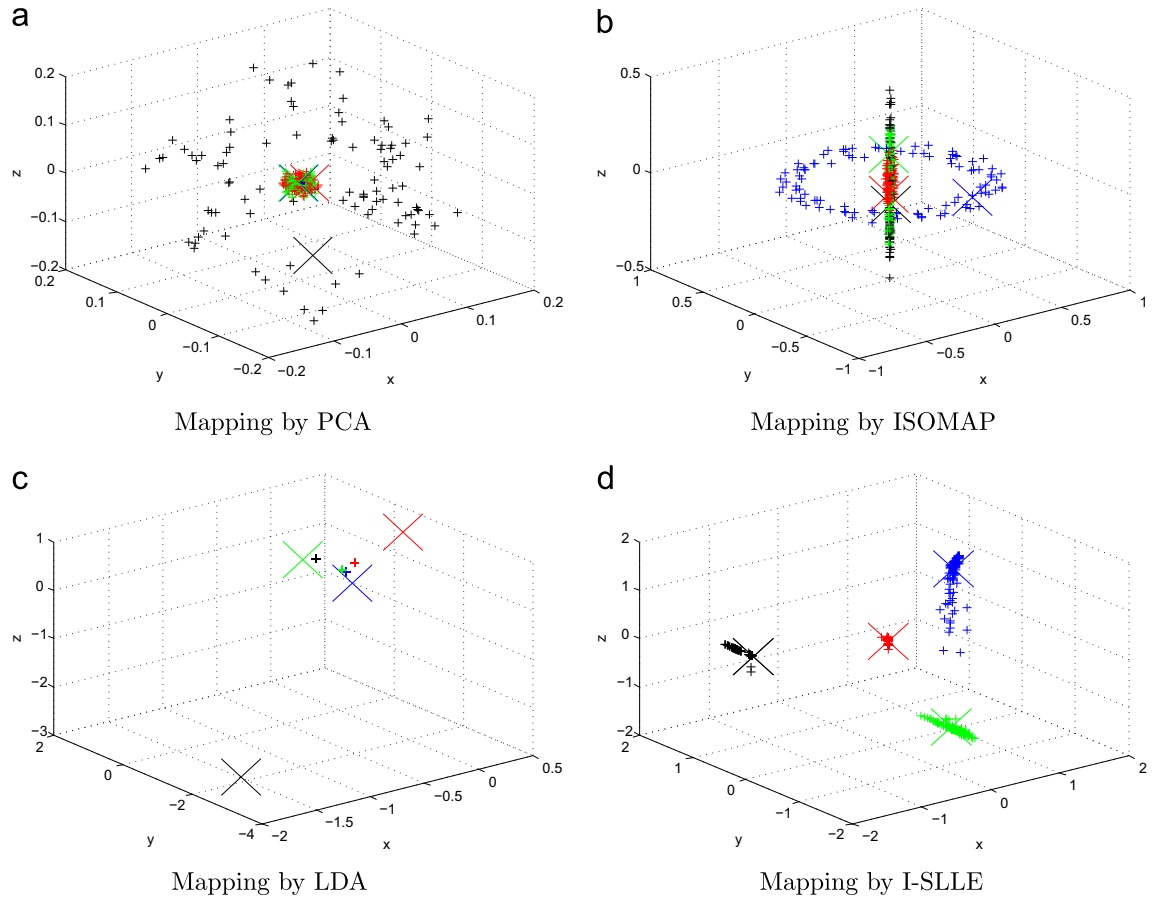


Fig. 6. Three-dimensional data of motor bearing. '+' denotes training data and 'X' represents test data. The blue points indicate normal data. The red points denote inner fault. The green points represent ball fault data. The black points indicate outer fault data. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

4. After dataset is updated, the residual variance f_2 will be calculate. If f_1 is larger than f_2, then the update will be accepted, otherwise it will be neglected.

5. SVM is trained in feature space, and the diagnostic result of the new samples is obtained by the trained SVM.

4. Experiment and analysis

In this section, to verify the effectiveness of I-SLLE for machinery diagnosis, synthetic and real datasets are used to test it. The specific parameters of all the datasets used in the experiments are summarized in Table 2. Firstly, I-LLE along with ILLE1, and ILLE3 are employed on synthetic datasets to prove the feasibility of I-LLE, and then the experiments are performed on three real datasets to examine our proposed method for machinery fault diagnosis.

4.1. Synthetic datasets

The three synthetic datasets include S-curve dataset, Punctured Sphere dataset, and Toroidal Helix dataset. All these datasets are standard benchmarks for dimension reduction algorithms. In this experiment, the three datasets are classified into training samples (1495 points) and test samples (5 points), and all the datasets are conducted as the same operation. At the beginning, the training dataset is mapped into a low-dimensional space by original LLE algorithm with $k=12$ and $d=2$; then, the test samples are calculated by ILLE1,LLE3, and I-LLE. The experimental results are shown in Fig. 2. As one can see, all the test datasets processed by the three incremental LLE are visually almost the same. Thus, when the size of test data is far less that of training data, the three algorithm has almost the same performance.

Table 3
The classification accuracy of each method to the bearing dataset with SVM classifier (%).

Test sample size per class	PCA	ISOMAP	LDA	I-SLLE
1	28	50	100	100
5	25	50	82	100
10	25	42	73	100
15	27	41	70	100
20	25	41	77	100
25	26	39	72	100
30	25	40	60	100

In actually, ILLE3 is the batch mode of ILLE1, hence, their results are equivalent. For quantitatively analyzing the performance of I-LLE, I-LLE along with I-LLE3 and original LLE are evaluated on the S-curve dataset with different size of test data. We randomly select 500 samples and 1000 samples from S-curve dataset as training data and test data, respectively. The parameters of the three algorithms are set as the same as the experiment mentioned above. We change the size of test data from 500 to 1000. As shown in Fig. 3, the comparison among ILLE3, original LLE and I-LLE is carried out in terms of error and computing time. The comparison results illustrate that ILLE3 shows a much large error with the size of test data increasing, while I-LLE and the original LLE display a similar accuracy, and the error is little relation to the size of test data. As for computing speed, ILLE3 is the fastest, I-LLE being nearly fast. Therefore, for large number of test data, I-LLE presents globally a better performance than the ILLE3 and the original LLE on S-curve dataset.

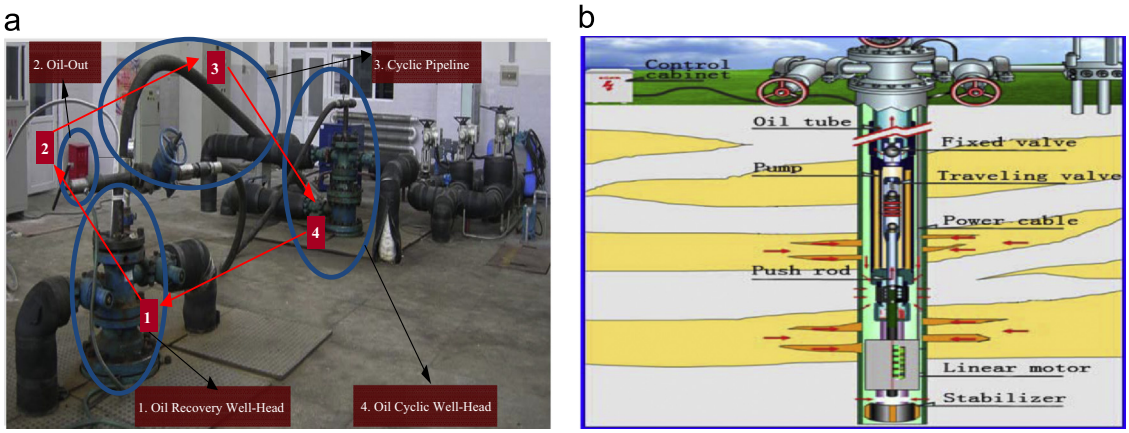
4.2. Iris dataset

Iris dataset is selected from the machine learning database UCI. There are three types of data in Iris dataset including Setosas data, Virginicas data, and Versicolors data. Each category contains 50 points, and each point has four features. A three-dimensional plot obtained by arbitrarily selecting three dimensions of Iris dataset is shown in Fig. 4(a). As shown in Fig. 4(a), one category is linearly separable, however, the other two categories are intertwined with each other, which are difficult to be linearly separated in the original space.

In order to illustrate the classification capacity of SLLE, four methods are considered as comparisons. The first method (PCA) is a global linear algorithm. The other three methods are ISOMAP, LTSA, and LLE, which are all local nonlinear methods. we visually compare the embedding results. As shown in Fig. 4, although the marginal separations of the different categories are improved by using PCA, ISOMAP, LTSA, and LLE, part of the data is still overlapping. It is quite clear that Iris dataset can be clearly separated by SLLE, hence, SLLE shows a perfect classification capability.

4.3. Motor bearing dataset

In this subsection, we evaluate the efficiency of I-SLLE by comparing it with PCA, ISOMAP, and LDA algorithms on a motor bearing dataset. PCA and LDA belong to linear methods, and they can construct an explicit mapping between the original space and the low-dimensional feature space, that is, they can easily deal



Real submersible plunger pump test platform

Internal structure chart of submersible plunger pump

Fig. 7. Submersible plunger pump test platform.

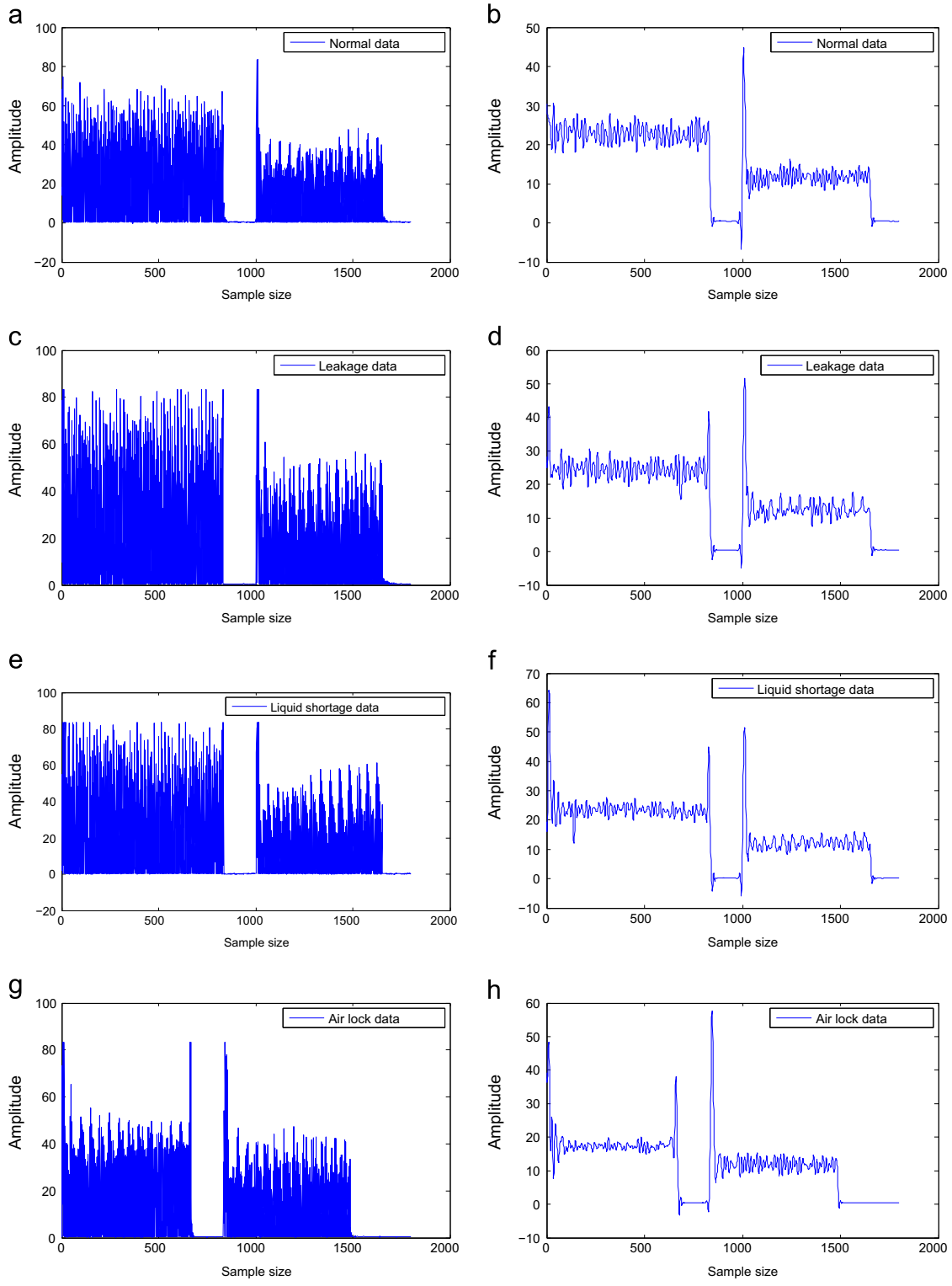


Fig. 8. Four typical time-domain signals of submersible plunger pump and their corresponding filtered signals.

with new samples. ISOMAP operates in a batch mode, it cannot deal with new samples directly. For obtaining an excellent result, we recompute ISOMAP to handle new samples. The motor bearing dataset is commonly utilized to validate fault diagnosis algorithms, and it has become a standard dataset for fault diagnosis. The dataset was sampled from a bearing test platform with the sampling frequency of 12 KHz and the motor speed of 1797 rpm. It includes one category normal data and three categories single fault dataset. These faults are generated by an electro-discharge machine with the fault fixing 0.007 in. (1 in.=25.4 mm) in diameter. The time-domain vibration signals of the normal and the fault data are shown in Fig. 5.

Randomly select 99 points and 1 point in each per class as training data and test data, respectively. First, we project the training data to the three-dimensional space; then, the new sample is handled by using the corresponding incremental

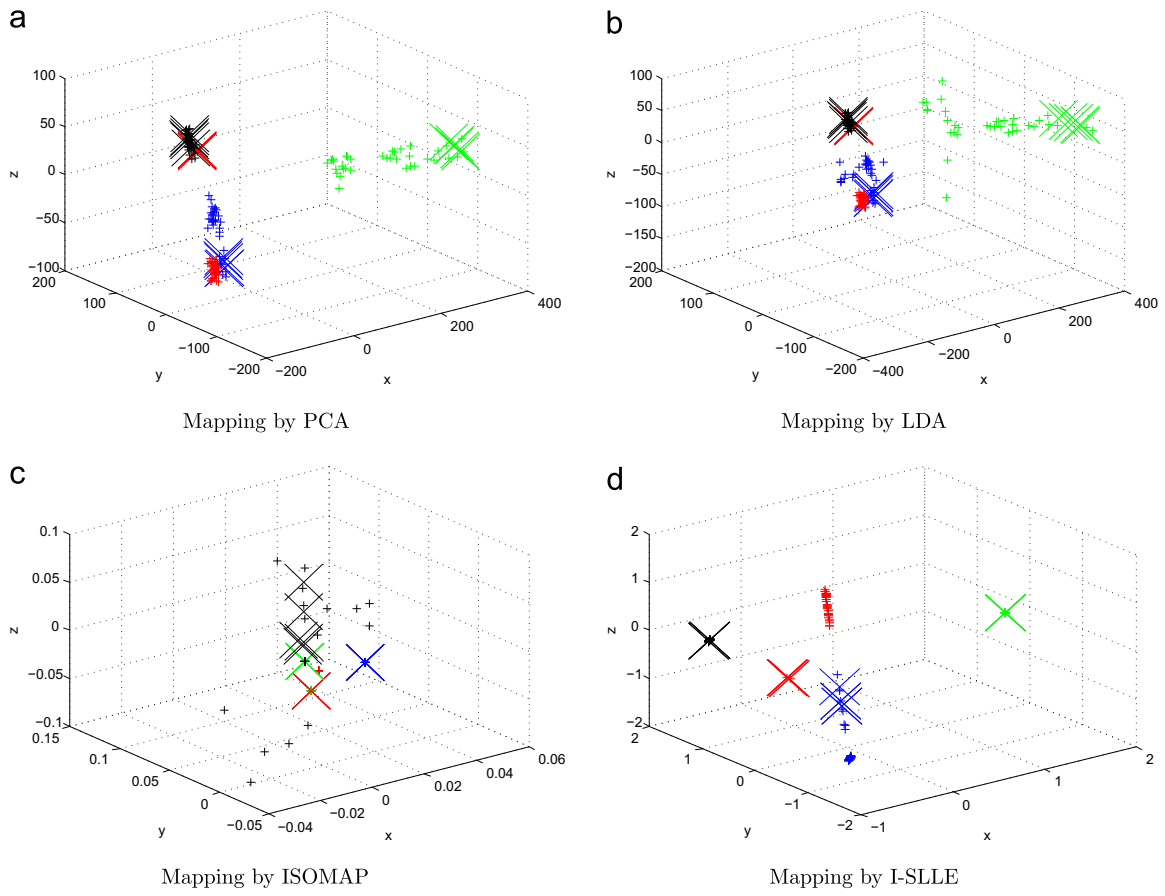


Fig. 9. Three-dimensional data of submersible plunger pump. '+' denotes training data and 'X' represents test data. The blue points indicate normal data. The red points denote leakage data. The green points represent liquid shortage data. The black points indicate air lock data. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

methods. For improving classification accuracy, we measure the similarity of the pairwise in frequency domain space. Fig. 6 shows the embedding results of training data and test data. In Fig. 6 (a) and (b), one can see that PCA and ISOMAP produced highly overlapping clusters, which are difficult to separate from each other, therefore, they have poor performance to classify this dataset. In the embedding result of LDA, the training data is completely separated, but LDA do fail for test data. This is because that LDA belongs to the linear method, when the dataset is very complex, LDA will fail. It is clear that the embedding result of SLLE spanning the embedded space possess sufficient discriminative power for data classification, and all the test samples are also classified correctly by using I-LLE algorithm. Furthermore, we conduct a quantitative comparison with the three competing algorithms by using SVM to calculate the recognition accuracy. We change the number of test data from 1 to 30 with an interval of five. As shown in Table 3, the result of PCA is the worst, and its highest accuracy is only 28%. The accuracies of LDA and ISOMAP are all low, while I-SLLE can recognize all the fault. Thus, I-SLLE outperforms the other three algorithms and it is suitable to employ I-SLLE to extract feature for fault diagnosis.

4.4. Submersible plunger pump fault dataset

Finally, we employ I-SLLE to extract the feature of submersible plunger pump for fault diagnosis. The submersible plunger pump fault dataset was sampled from an oil well platform. The platform and its diagrammatic sketch are shown in Fig. 7(a) and (b), respectively. The operation cycle of this equipment is 15 Hz, and we collect data with 500 Hz by an analog input module (NI PCI-6220). There are four kinds of data in this dataset including normal data, leakage data, liquid shortage data, and air lock data. The time-domain signals and their corresponding filtered signals are shown in Fig. 8. One can see that it is difficult to accurately identify the type of the data in the original domain. In this experiment, 140 training points are mapped into three-dimensional space by the PCA, LDA, ISOMA, and SLLE, respectively, and the other 20 test points are treated by their corresponding incremental methods. The three-dimensional plots are shown in Fig. 9. One can see that I-SLLE completely separates all the different classes data. Furthermore, SVM is employed in this feature space to calculate classification accuracy, and the results of the four algorithms are 36%, 70%, 93%, and 100%, respectively. Hence, it can be concluded that I-SLLE is effective for submersible plunger pump fault diagnosis.

5. Conclusions

In this paper, I-SLLE based fault diagnosis method is proposed. I-SLLE can deal with new samples and update old data, simultaneously. Since the existing incremental methods of LLE can only consider the new samples, I-SLLE provides a solution that overcomes this limitation. Furthermore, class information is added in the cost function of LLE, by which the structure information and the discrimination information are all contained in the embedding result. Hence, the embedding result is suitable for pattern recognition. The proposed method is performed on several synthetic and real world datasets, and the experimental results demonstrate the feasibility and the efficiency of our proposed method.

There are still several drawbacks existed in I-SLLE, which should be considered in future works. First, the parameters' selection of d, k, and α is a difficult task. Different parameters will bring significantly different results. Second, similarity measure method is a key factor for classification. An excellent measure criterion will enhance the classification accuracy, but a poor one may make algorithm invalid. Third, it is important for I-LLE to select a suitable iterative method. All of these issues should be deeply researched.

References

Ali, J., Saidi, L., Mouelhi, A., Chebel-Morello, B., Fnaiech, F., 2015. Linear feature selection and classification using pnn and sfam neural networks for a nearly online diagnosis of bearing naturally progressing degradations. Eng. Appl. Artif. Intell. 42, 67–81.

de Ridder, D., Kouropteva, O., Okun, O., Duin, R.P., 2003. Supervised locally linear embedding, Springer, Berlin, Heidelberg, pp. 333–341.

Echevarra, L., Santiago, O., Fajardo, J., Neto, A., 2014. A variant of the particle swarm optimization for the improvement of fault diagnosis in industrial systems via faults estimation. Eng. Appl. Artif. Intell. 28, 36–51.

Elgammal, A.M., Lee, C.-S., 2007. Nonlinear manifold learning for dynamic shape and dynamic appearance. Comput. Vis. Image Underst. 106 (1), 31–46.

Fernández, González, A.M., Díaz, J., Dorronsoro, J.R., 2015. Diffusion maps for dimensionality reduction and visualization of meteorological data. Neurocomputing 163, 25–37.

Ferracuti, F., Giantomassi, A., 2015. Electric motor defects diagnosis based on kernel density estimation and Kullback–Leibler divergence in quality control scenario. Eng. Appl. Artif. Intell. 44, 25–32.

Hu, F., Wang, C.T., Wu, Y.C., Fan, L.Z., 2014. Parameters selection of lle algorithm for classification tasks. In: Advanced Materials Research, vol. 1037. Trans Tech Publications, Switzerland, pp. 422–427.

Jin, X., Yuan, F., Chow, T., Zhao, M., 2014. Weighted local and global regressive mapping: a new manifold learning method for machine fault classification. Eng. Appl. Artif. Intell. 30, 118–128.

Kouropteva, O., Okun, O., 2002. Selection of the optimal parameter value for the locally linear embedding algorithm. In: 1st International Conference on Fuzzy Systems, pp. 359–363.

Kouropteva, O., Okun, O., Hadid, A., Soriano, M., 2005. Beyond locally linear embedding algorithm. Image Anal. 3540, 521–530.

Lawrence, K.S., Sam, T.R., 2003. Think globally, fit locally: unsupervised learning of nonlinear manifolds. J. Mach. Learn. Res. 4, 119–155.

Lee, J.H., McDonnell, K., 2014. A structure-based distance metric for high-dimensional space exploration with multidimensional scaling. IEEE Trans. Vis. Comput. Graph. 20 (3), 351–364.

Liu, Y., Yu, Z., Zeng, M., Wang, S., 2015. Dimension estimation using weighted correlation dimension method. Discrete Dyn. Nat. Soc. 114 (6), 1–10.

Roweis, S.T., Saul, L.K., 2000. Nonlinear dimensionality reduction by locally linear embedding. Science 290, 2323–2326.

Serdio, F., Lughofer, E., Pichler, K., Buchegger, T., Efendic, H., 2014a. Residual-based fault detection using soft computing techniques for condition monitoring at rolling mills. Inf. Sci. 259, 304–320.

Serdio, F., Lughofer, E., Pichler, K., Buchegger, T., Pichler, M., Efendic, H., 2014b. Fault detection in multi-sensor networks based on multivariate time-series models and orthogonal transformations. Inf. Fusion 20, 272–291.

Serdio, F., Lughofer, E., Pichler, K., Pichler, M., Buchegger, T., Efendic, H., 2015. Fuzzy fault isolation using gradient information and quality criteria from system identification models. Inf. Sci. 316, 18–39.

Shao, R., Hu, W., Wang, Y., Qi, X., 2014. The fault feature extraction and classification of gear using principal component analysis and kernel principal component analysis based on the wavelet packet transform. Measurement 54, 118–132.

Su, Z., Tang, B., Ma, J., Deng, L., 2014. Fault diagnosis method based on incremental enhanced supervised locally linear embedding and adaptive nearest neighbor classifier. Measurement 48, 136–148.

Tabatabaeipour, M., 2015. Active fault detection and isolation of discrete-time linear time-varying systems: a set-membership approach. Int. J. Syst. Sci. 46 (11), 1917–1933.

Tenenbaum, J.B., de Silva, V., Langford, J.C., 2000. A global geometric framework for nonlinear dimensionality reduction. Science 290 (5500), 2319–2323.

Wang, J., 2014. Real local-linearity preserving embedding. Neurocomputing 136, 7–13.

Wang, Z., Sun, X., 2013. Multiple kernel local fisher discriminant analysis for face recognition. Signal Process. 93 (6), 1496–1509.

Wang, S.B., Chen, X.F., Wang, Y., 2015. Nonlinear squeezing time-frequency transform for weak signal detection. Signal Process. 113, 195–210.

Yin, S., Huang, Z., 2015. Performance monitoring for vehicle suspension system via fuzzy positivistic c-means clustering based on accelerometer measurements. IEEE/ASME Trans. Mech. 20 (5), 2613–2620.

Yin, S., Ding, S.X., Xie, X., Luo, H., 2014. A review on basic data-driven approaches for industrial process monitoring. IEEE Trans. Ind. Electron. 61 (11), 6418–6428.

Yin, S., Li, X., Gao, H., Kaynak, O., 2015a. Data-based techniques focused on modern industry: an overview. IEEE Trans. Ind. Electron. 62 (1), 657–667.

Yin, S., Zhu, X., Kaynak, O., 2015b. Improved pls focused on key-performance-indicator-related fault diagnosis. IEEE Trans. Ind. Electron. 62 (3), 1651–1658.

Yoshua, B., Jean-Francois, P., Pascal, V., 2003. Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering, In Advances in Neural Information Processing Systems. MIT Press, Cambridge, MA, pp. 177–184.

Yu, J., 2013. A new fault diagnosis method of multimode processes using Bayesian inference based gaussian mixture contribution decomposition. Eng. Appl. Artif. Intell. 26 (1), 456–466.

Zhang, Z., Zha, H., 2002. Principal manifolds and nonlinear dimension reduction via local tangent space alignment. SIAM J. Sci. Comput. 26, 313–338.

Zhang, J., Shen, H., Zhou, Z., 2004. Unified locally linear embedding and linear discriminant analysis algorithm (ullelda) for face recognition. In: Advances in Biometric Personal Authentication. Lecture Notes in Computer Science, pp. 209–307.

Zhang, K., Jiang, B., Shi, P., 2015. Fault estimation observer design for discrete-time systems in finite-frequency domain. Int. J. Rob. Nonlinear Control 25 (9), 1379–1398.

Zhou, Changjun, Wang, Lan, Zhang, Qiang, Wei, Xiaopeng, 2014. Face recognition based on pca and logistic regression analysis. Opt.-Int. J. Light Electron Opt. 125 (20), 5916–5919.