Research Article

Kernel Local Linear Discriminate Method for Dimensionality Reduction and Its Application in Machinery Fault Diagnosis

Kunju Shi,¹ Shulin Liu,¹ Hongli Zhang,¹ and Bo Wang^{1,2}

¹ School of Mechatronics Engineering and Automation, Shanghai University, Shanghai 200072, China

² School of Mechanical and Electronic Engineering, Chuzhou University, Chuzhou 239000, China

Correspondence should be addressed to Shulin Liu; ls1346@126.com

Received 7 September 2013; Accepted 13 January 2014; Published 27 February 2014

Academic Editor: Gyuhae Park

Copyright © 2014 Kunju Shi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Dimensionality reduction is a crucial task in machinery fault diagnosis. Recently, as a popular dimensional reduction technology, manifold learning has been successfully used in many fields. However, most of these technologies are not suitable for the task, because they are unsupervised in nature and fail to discover the discriminate structure in the data. To overcome these weaknesses, kernel local linear discriminate (KLLD) algorithm is proposed. KLLD algorithm is a novel algorithm which combines the advantage of neighborhood preserving projections (NPP), Floyd, maximum margin criterion (MMC), and kernel trick. KLLD has four advantages. First of all, KLLD is a supervised dimension reduction method that can overcome the out-of-sample problems. Secondly, short-circuit problem can be avoided. Thirdly, KLLD algorithm can use between-class scatter matrix and inner-class scatter matrix more efficiently. Lastly, kernel trick is included in KLLD algorithm to find more precise solution. The main feature of the proposed method is that it attempts to both preserve the intrinsic neighborhood geometry of the increased data and exact the discriminate information. Experiments have been performed to evaluate the new method. The results show that KLLD has more benefits than traditional methods.

1. Introduction

With the information collection technology becoming more and more advanced, a huge number of data have been produced during mechanical equipment running process. The sensitive information which reflects the running status of the equipment has been submerged in a large amount of redundant data. Effective dimensionality reduction can solve this problem. Dimensionality reduction is one of the key technologies for equipment condition monitoring and fault diagnosis. Nonlinear and nonstationary vibration signals generated by the rolling bearing [1, 2] make the original high-dimensional feature space which consists of the statistical characteristics of the signal inseparable. The traditional linear dimensionality reduction methods such as PCA and ICA not only are under the assumption of global linear structure of the data but also use different linear transformation matrix to find the best low-dimensional projection. The classification information plays an important role. In nonlinear conditions such

as the original high dimensional feature space possesses a non-linear structure, however, the classification information is difficult to obtain by linear methods. KPCA is a traditional nonlinear dimensionality reduction method, which achieves the task of dimensionality reduction by discarding relatively small projection in a higher-dimensional linear space. In addition, KPCA aims to find the principal components with the largest variance, which may cause the loss of useful discriminate information [3].

Manifold learning is a data-driven approach and can reveal the underlying nature of the complex data structure, which provides a new approach for the analysis of the intrinsic dimension based on the data distribution. Manifold learning has got a series of research achievements in the feature extraction [4–6]. Actually, manifold learning method falls broadly into two categories [7] which have different advantages and disadvantages: global (Isomap [8]) and local (locally linear embedding [9]). In [10], the author points out that, as for the discriminate analysis, the local structure

is usually more important than the global structure when there are not enough samples. As for local manifold learning, local linear embedding (LLE) is an algorithm which has many advantages such as global optimal solution and fast calculation. Furthermore, its minimum reconstruction error weights can keep data local neighborhood geometric properties unchanged when data exhibition shrinks and rotates. So LLE algorithm is applied to the fault feature extraction [11–15].

NPP [16] is one of the manifold learning methods, whose central idea is based on LLE by introducing a linear transform matrix. NPP has been successfully applied in famous “Swiss roll” and “S-curve” dataset dimension reduction. The algorithm assumes that the structure of data on the local significance is linear. However, when the data manifold has a larger bending, manifold learning method will result in the short-circuit problem.

As for these issues, a fault feature extraction method named KLLD is proposed in the paper. This method studies both the iris dataset and the rolling bearing original feature dataset constructed by wavelet packet energy with dimensionality reduction application. The effectiveness of this method is verified by contrast with conventional analysis methods.

The rest of this paper is organized as follows. In Section 2, we review briefly the LLE, NPP, Floyd, and MMC algorithm. In Section 3, firstly, KLLD algorithm proposed in this paper is deduced; secondly, the short-circuit problem is introduced and Floyd algorithm is employed to overcome this drawback; lastly, based on the LLD, KLLD, and Floyd algorithm, the calculation steps of LLD and KLLD are designed. In Section 4, we design a KLLD experiment process for the dimension reduction of iris and rolling bearing database; then we apply the KPCA, NPP, LLD, and KLLD algorithm to dataset dimension reduction. Conclusions are made and several issues for future study are addressed in Section 5.

2. Basic Principle

2.1. Local Linear Embedding. The LLE (local linear embedding) algorithm is a new nonlinear data dimension reduction technology, which utilizes local neighborhood relation to learn the global structure of nonlinear manifolds.

For the given N D -dimensional real-valued vectors $\mathbf{x}_i \in \mathbf{X}$ ($i = 1, \dots, N$), assume that the vectors of each point and its nearest neighbor lie in the local linear space, by weighting coefficients $\mathbf{W}_{ij}$ ($i = 1, \dots, N; j = 1, \dots, k$) with k neighborhood which belongs to $\mathbf{x}_i$ to reconstruct $\mathbf{x}_i$. $\mathbf{W}_{ij}$ is selected by minimizing the cost function. That is,

$$\min J_1(\mathbf{W}) = \sum_{i=1}^n \left\| \mathbf{x}_i - \sum_{j=1}^k w_{ij} \mathbf{x}_{ij} \right\|. \quad (1)$$

In LLE algorithm low dimensional points can be reconstructed by high-dimensional matrix. Each of the high-dimensional data $\mathbf{x}(:, j)$ can be obtained by $\mathbf{M}$ eigenvalue decomposition of formula (2) and then identify the bottom $d + 1$ eigenvectors corresponding to its smallest $d + 1$ eigenvalues. Then $\mathbf{Y}$ is a matrix that is constructed by using these

eigenvectors and discarding the eigenvectors corresponding to its smallest eigenvalues of the matrix. Derivation can be found in the literature [9]:

$$\mathbf{M} = (\mathbf{I} - \mathbf{W})^T \times (\mathbf{I} - \mathbf{W}). \quad (2)$$

In summary, we state LLE as the following algorithm.

Step 1. Select neighbors by K nearest neighbor algorithm.

Step 2. Reconstruct with linear weights by (1).

Step 3. Map to embed coordinates $\mathbf{Y}$.

2.2. Neighborhood Preserving Projections. LLE algorithm processes so many advantages; however, its computation cost is more expensive compared to other linear dimensionality reduction methods. Moreover, it does not have the ability to map new test datasets into low-dimension space directly, which is called out-of-sample problems. NPP derives from LLE and inherits LLE’s neighborhood property. However, it modifies the LLE by introducing a linear matrix $\mathbf{A}$. NPP can overcome the drawbacks of LLE.

The low-dimensional feature datasets $\mathbf{Y}$ ($\mathbf{Y} \in R^d$) are obtained from original datasets $\mathbf{X}$ ($\mathbf{X} \in R^D$) by the projection matrix $\mathbf{A}$ ($\mathbf{Y} = \mathbf{A}^T \mathbf{X}$). The main idea of the method is to compute the transform matrix $\mathbf{A}$ on the premise of keeping the smallest reconstruction error (see (3) and (4)):

$$\min J_2(\mathbf{Y}) = \sum_{i=1}^N \left\| \mathbf{y}_i - \sum_{j=1}^k w_{ij} \mathbf{y}_j \right\| = \text{trace}(\mathbf{Y} \mathbf{M} \mathbf{Y}^T) \quad (3)$$

$$\text{s.t.} \quad \frac{1}{N-1} \mathbf{Y} \mathbf{Y}^T = \mathbf{I}. \quad (4)$$

Let $\mathbf{Y} = \mathbf{A}^T \mathbf{X}$. (3) can be transformed to the following:

$$\begin{aligned} \min J_2(\mathbf{Y}) &= \text{trace}((\mathbf{A}^T \mathbf{X}) \mathbf{M} (\mathbf{A}^T \mathbf{X})^T) \\ &= \text{trace}(\mathbf{A}^T (\mathbf{X} \mathbf{M} \mathbf{X}^T) \mathbf{A}), \end{aligned} \quad (5)$$

where $\mathbf{M}$ is obtained by solving (2). The Lagrange extreme method is used:

$$((\mathbf{X} \mathbf{X}^T)^{-1} (\mathbf{X} \mathbf{M} \mathbf{X}^T)) \mathbf{A}^T = \lambda \mathbf{A}^T. \quad (6)$$

Obviously, we know the projection matrix $\mathbf{A}$ is generalized eigenvectors of $((\mathbf{X} \mathbf{X}^T)^{-1} (\mathbf{X} \mathbf{M} \mathbf{X}^T))$.

2.3. Floyd Algorithm. This algorithm was proposed by Floyd, which is used to calculate the shortest distance between two specified points. $\mathbf{D}_m$ is a matrix of $N \times N$ and $i, j \in \{1, 2, \dots, N\}$ are specific numbers for each point in $\mathbf{D}_m$. d_{ij}^m represents the shortest path between vertex i and vertex j . Point $1, 2, \dots, m - 1$ is an intermediate point; the Floyd algorithm contains two calculation steps.

Step 1. Initialization: computing d_{ij}^0 . We set $d_{ij}^0 = \infty$ when such shortest path does not exist; otherwise $d_{ij}^0 = 0$.

Step 2. For $m = 1, 2, \dots, N$, using (7) to loop iteration, calculate short path of each point in $\mathbf{D}_m$:

$$d_{ij}^m = \min \{d_{im}^{m-1} + d_{mj}^{m-1}, d_{ij}^{m-1}\}. \quad (7)$$

2.4. Maximum Margin Criterion. LDA (linear discriminate analysis) is a popular linear feature extractor. The key step is to find a transform matrix $\mathbf{Q} \in R^{D \times d}$ under the condition of Fisher criterion maximized, where D is the number of dimension in original dataset and $\mathbf{X}$ and d are the number of $\mathbf{Y}$'s dimension [17]:

$$J_F(\mathbf{Q}) = \frac{\mathbf{Q}^T \mathbf{S}_b \mathbf{Q}}{\mathbf{Q}^T \mathbf{S}_w \mathbf{Q}}, \quad (8)$$

where

$$\begin{aligned} \mathbf{S}_b &= \sum_{i=1}^c p_i (m_i - m)(m_i - m)^T, \\ \mathbf{S}_w &= \sum_{i=1}^c p_i \mathbf{S}_i. \end{aligned} \quad (9)$$

$\mathbf{S}_b$ represents the distance matrix between classes and $\mathbf{S}_w$ represents distance matrix within class. c is the number of classes; m_i and p_i are the mean vector and a priori probability of class i , respectively. m is overall mean vector.

However, we can find drawbacks, because (8) cannot be applied when $\mathbf{S}_w$ is singular due to the small sample size problem. MMC method has been proposed in the literature [17] to overcome these drawbacks. According to LDA, MMC can be represented in the following:

$$J = \text{tr}(\mathbf{S}_b - \mathbf{S}_w). \quad (10)$$

3. Kernel Local Linear Discriminate Algorithm

3.1. Local Linear Discriminate Algorithm. According to LLT-STA derivation process in the literature [18], the basic idea of the LLD proposed in this paper is that if the linear transform matrix $\mathbf{A}$ of NPP (see (5)) satisfies (10), the ability of discriminate different class of data will be greatly improved. This problem can be expressed as a multiobjective optimization problem:

$$\begin{aligned} \min_{\mathbf{A}} \quad & \text{tr} \{ \mathbf{A}^T (\mathbf{X} \mathbf{M} \mathbf{X}^T) \mathbf{A} \} \\ \max_{\mathbf{A}} \quad & \text{tr} \{ \mathbf{A}^T (\mathbf{S}_b - \mathbf{S}_w) \mathbf{A} \} \\ \text{s.t.} \quad & \frac{1}{N-1} \mathbf{Y} \mathbf{Y}^T = \mathbf{I}. \end{aligned} \quad (11)$$

Equation (11) can be changed into constrained optimization problem:

$$\begin{aligned} \min_{\mathbf{A}} \quad & \text{tr} \{ \mathbf{A}^T [(\mathbf{X} \mathbf{M} \mathbf{X}^T) - (\mathbf{S}_b - \mathbf{S}_w)] \mathbf{A} \} \\ \text{s.t.} \quad & \frac{1}{N-1} \mathbf{A}^T (\mathbf{X} \mathbf{X}^T) \mathbf{A} = \mathbf{I}. \end{aligned} \quad (12)$$

FIGURE 1: The short-circuit problem.

TABLE 1: Comparison of four dimensionality reduction methods for iris dataset classification.

	<i>I. setosa</i>	<i>I. versicolor</i>	<i>I. virginica</i>	Average accuracy
KPCA	0.960	0.760	0.760	0.8267
NPP	1.0	0.80	1.0	0.9333
LLD	1.0	0.920	0.920	0.9467
KLLD	1.0	0.920	1.0	0.9733

Lagrange multiplier method is used to solve this problem; that is,

$$\begin{aligned} \frac{\partial}{\partial \mathbf{A}} \left\{ \mathbf{A}^T [(\mathbf{X} \mathbf{M} \mathbf{X}^T) - (\mathbf{S}_b - \mathbf{S}_w)] \mathbf{A} \right. \\ \left. - \frac{\lambda}{N-1} \mathbf{A}^T (\mathbf{X} \mathbf{X}^T) \mathbf{A} + \lambda \mathbf{I} \right\} = 0 \end{aligned} \quad (13)$$

further deduced as follows:

$$[(\mathbf{X} \mathbf{M} \mathbf{X}^T) - (\mathbf{S}_b - \mathbf{S}_w)] \mathbf{A} = \frac{\lambda}{N-1} (\mathbf{X} \mathbf{X}^T) \mathbf{A}. \quad (14)$$

Equation (14) is converted into

$$(N-1)(\mathbf{X} \mathbf{X}^T)^{-1} [(\mathbf{X} \mathbf{M} \mathbf{X}^T) - (\mathbf{S}_b - \mathbf{S}_w)] \mathbf{A} = \lambda \mathbf{A}; \quad (15)$$

then we can find the projection matrix $\mathbf{A}$ is eigenvectors of $(N-1)(\mathbf{X} \mathbf{X}^T)^{-1} [(\mathbf{X} \mathbf{M} \mathbf{X}^T) - (\mathbf{S}_b - \mathbf{S}_w)]$.

3.2. Kernel Local Linear Discriminate Algorithm. Suppose that φ is a nonlinear mapping to some feature space $\mathcal{F}$; (15) can be changed into the following:

$$\begin{aligned} (N-1) [(\varphi(\mathbf{X}) \mathbf{M} \varphi(\mathbf{X})^T) - (\mathbf{S}_b^\varphi - \mathbf{S}_w^\varphi)] \mathbf{A}^\varphi \\ = \lambda \varphi(\mathbf{X}) (\varphi(\mathbf{X})^T) \mathbf{A}^\varphi. \end{aligned} \quad (16)$$

To find local linear discrimination information in the feature space $\mathcal{F}$, we need (16) dot product form of input patterns. Then we replace the dot product form with one of the kernel functions. From the theory of reproducing kernels we know that any solution $\mathbf{A}^\varphi \in \mathcal{F}$ must lie in the span of all

FIGURE 2: Scatter plots of iris dataset by different features: (a) KPCA Matrix feature of dataset1, (b) KPCA Matrix feature of dataset2, (c) NPP Manifold feature of dataset1, (d) NPP Manifold feature of dataset2, (e) LLD Manifold feature of dataset1, (f) LLD Manifold feature of dataset2, (g) KLLD Manifold feature of dataset1, and (h) KLLD Manifold feature of dataset2.

TABLE 2: Original features constructed by wavelet packet energy spectrums of rolling bear dataset1.

Energy band	Normal				Inner race fault					Outer race fault			
	x_1	x_2	$\cdots$	x_{40}	x_{41}	x_{42}	$\cdots \cdots \cdots$	x_{80}	$\cdots$	x_{121}	x_{122}	$\cdots$	x_{160}
1	0.6363	0.6593	$\cdots$	0.6046	0.0172	0.0181	$\cdots$	0.0160	$\cdots$	0.0303	0.0287	$\cdots$	0.0284
2	0.1865	0.1622	$\cdots$	0.2291	0.4220	0.4204	$\cdots$	0.4222	$\cdots$	0.1484	0.1579	$\cdots$	0.2012
3	0.0074	0.0075	$\cdots$	0.0071	0.0648	0.0701	$\cdots$	0.0742	$\cdots$	0.1505	0.1522	$\cdots$	0.1463
4	0.1654	0.1666	$\cdots$	0.1554	0.0451	0.0373	$\cdots$	0.0338	$\cdots$	0.0220	0.0223	$\cdots$	0.0275
5	0.0004	0.0004	$\cdots$	0.0004	0.0107	0.0110	$\cdots$	0.0114	$\cdots$	0.0010	0.0010	$\cdots$	0.0009
6	0.0013	0.0012	$\cdots$	0.0012	0.0311	0.0290	$\cdots$	0.0256	$\cdots$	0.0085	0.0087	$\cdots$	0.0073
7	0.0005	0.0004	$\cdots$	0.0005	0.3345	0.3383	$\cdots$	0.3387	$\cdots$	0.6169	0.6076	$\cdots$	0.5681
8	0.0023	0.0025	$\cdots$	0.0018	0.0745	0.0758	$\cdots$	0.0782	$\cdots$	0.0225	0.0216	$\cdots$	0.0203

TABLE 3: Distances between each point of rolling bear dataset1.

	x_1	x_2	$\dots$	x_{40}	x_{41}	x_{42}	$\dots$	x_{80}	$\dots$	x_{121}	x_{122}	$\dots$	x_{160}
x_1	0.0000	0.0062	$\dots$	0.0206	3.9698	3.9639	$\dots$	3.8892	$\dots$	3.3080	303006	$\dots$	3.0511
x_2	0.0062	0.0000	$\dots$	0.0482	4.1455	4.1385	$\dots$	4.0637	$\dots$	3.3738	3.3704	$\dots$	3.1379
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{40}	0.0206	0.0482	$\dots$	0.0000	3.6692	3.6591	$\dots$	3.5795	$\dots$	3.1847	3.1706	$\dots$	2.8929
x_{41}	3.9698	4.1455	$\dots$	3.6692	0.0000	0.0083	$\dots$	0.0442	$\dots$	2.3464	2.2919	$\dots$	2.1105
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{80}	3.8892	4.0637	$\dots$	3.5795	0.0442	0.0153	$\dots$	0.0000	$\dots$	2.1124	2.0596	$\dots$	1.8712
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{160}	3.0511	3.1379	$\dots$	2.8929	2.1105	1.9920	$\dots$	1.8712	$\dots$	0.0322	0.0239	$\dots$	0.0000

TABLE 4: Comparison of within-class distance of low dimension using four dimensionality reduction methods.

	Normal	Inner race fault	Ball fault	Outer race fault
KPCA	1.62×10^3	1.51×10^6	5.74×10^6	6.76×10^7
	2.67×10^7	1.63×10^7	4.97×10^6	1.33×10^8
NPP	5.14×10^{-4}	5.03×10^{-4}	1.42×10^{-3}	1.68×10^{-3}
	7.01×10^{-4}	2.64×10^{-4}	2.08×10^{-4}	1.53×10^{-4}
LLD	1.84×10^{-3}	5.98×10^{-4}	1.81×10^{-3}	3.28×10^{-3}
	2.91×10^{-3}	1.61×10^{-3}	9.31×10^{-4}	7.58×10^{-4}
KLLD	1.28×10^{-13}	7.80×10^{-12}	5.38×10^{-12}	2.09×10^{-12}
	2.46×10^{-3}	1.28×10^{-3}	7.06×10^{-4}	5.87×10^{-4}

TABLE 5: Ratio of between-class distance and within-class distance using four dimensionality reduction methods.

	KPCA	NPP	LLD	KLLD
Dataset1	159.5	517.7	226.6	4.38×10^{11}
Dataset2	38.97	47.28	120.8	186.4

training samples in $\mathcal{F}$. Therefore, expansion terms for $\mathbf{A}^\varphi$ can be written as follows:

$$\mathbf{A}^\varphi = \sum_{l=1}^N \beta_l \phi(x_l). \quad (17)$$

Combining (16) and (17) and then multiplying $[\phi(\mathbf{X}_p)]^T$ at both sides of (16), we have

$$\begin{aligned} & (N-1) \varphi(\mathbf{X}_p)^T \left[\varphi(\mathbf{X}) \mathbf{M} \varphi(\mathbf{X})^T - (\mathbf{S}_b^\varphi - \mathbf{S}_w^\varphi) \right] \\ & \cdot \left[\sum_{l=1}^N \beta_l \varphi(x_l) \right] \\ & = \lambda \varphi(\mathbf{X}_p)^T \varphi(\mathbf{X}) \varphi(\mathbf{X})^T \left[\sum_{l=1}^N \beta_l \varphi(x_l) \right]. \end{aligned} \quad (18)$$

TABLE 6: Comparison of four dimensionality reduction methods for classification.

	Normal	Inner race fault	Ball fault	Outer race fault	Average accuracy
KPCA	1.0	1.0	0.15	1.0	0.7875
NPP	1.0	1.0	0.775	1.0	0.9438
LLD	1.0	1.0	0.90	1.0	0.975
KLLD	1.0	1.0	1.0	1.0	1.0

FIGURE 3: Scheme of KLLD dimension reduction for rolling bear dataset.

We set $K(x_i, x_j) = (\phi(x_i) \cdot \phi(x_j))$, and then (18) can be rewritten as follows:

$$\begin{aligned}
 (N-1) \left[\mathbf{KMK} - \sum_{i=1}^C \left(\left(\frac{1}{n_i} \right) \left(\sum_{j=1}^{n_i} \mathbf{K}_{i,j} - \sum_{h=1}^n \mathbf{K}_{i,h} \right) \right. \right. \\
 \left. \left. - \sum_{l=1}^D \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{K}_{i,j} - \frac{1}{N} \sum_{k=1}^D \mathbf{K}_{l,k} \right) \right) \right] \beta \\
 = \lambda \mathbf{KK} \beta,
 \end{aligned} \tag{19}$$

where n_i is the number of i th samples, C is the number of samples, and D is the number of sample dimensions. Taking polynomial kernel function ($k(x, y) = \phi(x) \cdot \phi(y) = (x \cdot y + 1)^q$) as the sample, $\mathbf{K}_{i,j}$ can be rewritten as $\mathbf{K}_{i,j} = (x_i \cdot x_j + 1)^q$. Similar to Section 3.1, (19) can be considered as a generalized eigenvalue decomposition problem.

3.3. Short-Circuit Problem. Traditional Euclidean distance method has many advantages such as perceptual intuitional, easy to understand and calculation. However, Euclidean method could easily lead to short-circuit problem [19], when the high-dimension space possesses a larger hypersurface curvature. Short-circuit problem refers to the fact that a point's neighbor mixed with different types of points, which results in discrimination information cannot be extracted effectively. Distribution of the two-type data in the two-dimensional space is shown in Figure 1.

Figure 1 shows two types of points including round and square. Under the condition of Euclidean distances, round12's five close neighbors are {round11, round13, square2, square3, and square4}. This phenomenon will lead to distortion of data dimensionality reduction in low-dimensional space. In order to overcome this drawbacks, we created a connect graph. The point of different type is not connected and we deem the distance between the unconnected points is infinity. In this way the round12's five nearest neighbor points are {round9, round10, round11, round13, and round14}. Therefore, the Floyd algorithm in Section 2.3 is used to find the distance between points in the figure after establishing a connection diagram in high-dimensional sample space. To find the right nearest neighbor point in the LLD algorithm, using the Floyd algorithm can effectively avoid the problem of mixing different types of data samples.

3.4. Steps of LLD Calculation. According to Sections 3.1, 3.2, and 3.3 analysis, we state LLD as the following algorithm.

Input. One has the original space $D \times N$ matrix $\mathbf{X}$, close neighbor points K , connection distance c , low-dimensional embedding dimension d ($d < D$), and Kernel parameter q .

Output. One has $d \times N$ low-dimensional matrix $\mathbf{Y}$.

Step 1. This step involves computing the Euclidean distance between any two points $\text{dis } \mathbf{x}(i, j)$ ($i = 1, 2, \dots, N, j = 1, 2, \dots, K$).

Step 2. Set connection threshold c , determine points value, and construct a weighted graph similar to Figure 1.

Step 3. The distance between $\mathbf{x}_i$ and the rest points is calculated. $\mathbf{x}_i$'s K nearest neighborhoods are selected under the condition that minimum distance is determined according to Section 2.3.

Step 4. Reconstruct weighting matrix $\mathbf{W}$, which is calculated according to formula (1).

FIGURE 4: Waveform of vibration signals from bearing under four conditions: (a) normal, (b) inner race fault, (c) ball fault, and (d) outer race fault.

FIGURE 5: Wavelet packet energy spectrums of bearing vibration signals under four conditions: (a) normal, (b) inner race fault, (c) ball fault, and (d) outer race fault.

FIGURE 6: Finding optimal K using residual variance.

Step 5. Calculate matrix $\mathbf{A}$ according to formula (15) and $(N-1)(\mathbf{X}\mathbf{X}^T)^{-1}[(\mathbf{X}\mathbf{M}\mathbf{X}^T) - (\mathbf{S}_b - \mathbf{S}_w)]$ eigenvalue decomposition matrix, and then find the low-dimensional embedding $d+1$ smallest eigenvalue, 2 to $d+1$ smallest eigenvalue corresponding feature vector $\mathbf{A}$.

Step 6. Calculate $\mathbf{Y}$ by $\mathbf{Y} = \mathbf{A}^T \mathbf{X}$.

3.5. Steps of KLLD Calculation. Besides the input and output of LLD, kernel parameter q also should be considered as for KLLD's input. There are 3 calculation steps.

Step 1. Reconstruct weighting matrix $\mathbf{W}$ as Step 1 to Step 4 of Section 3.4, and then calculate $\mathbf{M}$ according to formula (2).

Step 2. Calculate β according to formula (19); then $\mathbf{A}^\varphi$ can be obtained by (17).

Step 3. Calculate $\mathbf{Y}$ by $\mathbf{Y} = (\mathbf{A}^\varphi)^T \mathbf{X}$.

4. Application Analysis

4.1. Iris Dataset Dimensionality Reduction. We evaluated the performance of the new approach on the iris plants database. *I. setosa*, *versicolor*, and *virginica* are included in this dataset. Sepal length, sepal width, petal length, and petal width are the characteristics of the plant samples. The number of each of the plant samples is 50. We divided them into two parts equally named dataset1 and dataset2. So there were 25 plant samples in each class of the new database. KPCA, NPP, and LLD were also used in this section to demonstrate the advantage of the dimensionality reduction method. Polynomial kernel function whose parameter $q = 25$ was employed in KLLD and KPCA. The number of close neighbor points is $K = 14$, which was used in NPP, LLD, and KLLD method. The results of iris database dimension reduction are shown in Figure 2.

KPCA and NPP methods hardly discriminate three types of the plant as shown in Figures 2(a)–2(d), because of

the fact that both of them reduce dimension for describing data. That is to say, they keep the information not discarded during dataset dimension reduction. Figures 2(e)–2(h) show the results of LLD and KLLD methods. There are points representing different plant overlaps in Figures 2(e)–2(f), because LLD is a linear method. As shown in Figures 2(g)–2(h), the KLLD method can discriminate different kinds of plants properly. In order to investigate the accuracy of classification, SVM [20] is used as the classifier. Table 1 shows the accuracy of SVM classification. Dataset1 is the training dataset and dataset2 is the testing dataset. Table 1 illustrates that the KLLD method obtains the highest classification accuracy.

4.2. Rolling Bear Fault Datasets Dimensionality Reduction Experiment

4.2.1. KLLD Experiment Process Design. The calculation steps are designed as follows.

Step 1. Collect vibration signal of the rolling bear.

Step 2. Wavelet packet energy is used to construct the original features datasets.

Step 3. Projection matrix $(\mathbf{A}^\varphi)^T$ is obtained by KLLD.

Step 4. Find out dimension reduction result by $\mathbf{Y} = (\mathbf{A}^\varphi)^T \mathbf{X}$.

KLLD dimensionality reduction process for rolling bearing fault datasets is shown in Figure 3.

4.2.2. Wavelet Packet Energy Original Feature Construction. Under normal and fault operating conditions, time-domain waveform signals are shown in Figure 4.

The time-domain waveform characteristic of bearing inner race fault is typical shock component. The waveform of normal bearing rolling shows the feature of stable and little fluctuation in amplitude. The waveform of rolling element bearings fault includes random single punch strike component, while the time-domain waveform of bearing outer race fault is very similar to the inner race fault waveform. It is hard to grasp the rolling bear feature of different fault condition only from time-domain waveform. Wavelet packet analysis is a precise method for signal analysis. It is widely used in bearing fault diagnosis currently. So we use this method to construct the original feature. Typical fault of wavelet packet energy is shown in Figure 5. We can find that the different rolling bear faults signals which are processed by the wavelet packet decomposition have significantly different amplitude in different frequency bands of energy.

4.2.3. Wavelet Packet Energy Original Feature Construction. To verify the validity of the KLLD method, the experiment was performed on Electrical Engineering Laboratory rolling bear vibration database of Case Western Reserve University. We selected the bearing model SKF6203, with the running speed 1730 rpm under normal, inner race fault, ball fault,

FIGURE 7: Scatter plots of bearing vibration samples by different features: (a) KPCA Matrix feature of dataset1, (b) KPCA Matrix feature of dataset2, (c) NPP Manifold feature of dataset1, (d) NPP Manifold feature of dataset2, (e) LLD Manifold feature of dataset1, (f) LLD Manifold feature of dataset2, (g) KLLD Manifold feature of dataset1, and (h) KLLD Manifold feature of dataset2.

and outer race fault. They were processed by wavelet packet decomposition and two original feature datasets were constructed, named dataset1 and dataset2. Both of the datasets have 40 points. Table 2 shows the original features of dataset1.

4.2.4. Parameter Settings. According to [21], the authors calculated the optimal embedding dimension of manifold learning algorithm by the following:

$$d = q - 1, \quad (20)$$

where d is the optimal embedding dimension and q is the number of categories. This study considers four operational states of the rolling bear, so the low-dimensional space is 3 by (20). In addition, the distribution of the data can be clearly illustrated in 3-dimensional spaces.

The parameter K is the quantity of nearest neighbor. It is one of the most important parameters in manifold learning algorithm, because if the number of nearest neighbors is too large the small-scale structure of the manifold could be eliminated and the whole manifold would be smooth. On the contrary, if the number of nearest neighborhoods is too little, the successive manifold may be divided into disjointed submanifolds. Residual variance can be defined as $1 - \rho_{D_x D_y}^2$, where D_x, D_y represent the Euclidian distance matrixes of each point in $\mathbf{X}$ and $\mathbf{Y}$, respectively. ρ represents standard linear correlation. The smaller the value of residual variance is, the better high-dimension dataset can be embedded into low dimension [9]. The optimal value of K can be found by

$$K_{\text{opt}} = \arg \min_K \left(1 - \rho_{D_x D_y}^2 \right). \quad (21)$$

LLE is the basic version of NPP, LLD, and KLLD and it can be used to determine the optimal K of datasets. Dataset1 is input $\mathbf{X}$ of LLE algorithm, the value of K is from 2 to 39, and the results are illustrated in Figure 6. So $K = 10$ is the optimal quantity of neighborhood.

We can find that the distance between different classes is large in numerical while that within the same class is small from Table 3. In order to guarantee that the number of neighborhood in the same class is 10, the connection distance is $c = 0.441$ according to Table 3. The number of neighbors is $K = 10$. Polynomial kernel function is used in KLLD and KPCA and its parameter $q = 35$. Dataset2 is also done the same as we do in dataset1. The KLLD algorithm is used to calculate dataset1 and dataset2.

4.2.5. Calculation and Discussion. KPCA, NPP, and LLD algorithms are also used to analyze the effectiveness of KLLD. Table 2 is input matrix. Figure 7 shows the results of distribution in three-dimensional space.

As shown in Figures 7(a)-7(b), inner race fault and ball fault overlap with each other. KPCA hardly distinguishes different class of points during dimension reduction. Figures 7(c)-7(d) illustrate that normal and inner race fault can be distinguished; however, ball fault and outer race fault have some slight aliasing, especially in dataset2. We can find the same phenomenon in Figures 7(e)-7(f). The results

of KLLD are directly shown in 7(c)-7(d). KLLD algorithm can distinguish different classes of rolling bearing dataset. Figure 7 suggests that the different states of discriminating sensitive characteristics are retained.

The inner-class distance in low dimension is calculated and it is shown in Table 4, in order to evaluate the effectiveness of the KLLD method precisely. It shows that the proposed method has better clustering ability. To quantitatively evaluate the separability of the method, the sample ratio of between-class average distance and average intraclass distance is calculated. We can find in Table 5 that KLLD is 4.38×10^{11} while other methods are less than 517.7 in dataset1, and KLLD is 186.4 while other methods are less than 120.8. SVM has been used to calculate the classification accuracy of low-dimension dataset. The results are shown in Table 6, which illustrate that KLLD-SVM method can recognize each condition of rolling bear vibration signal.

5. Conclusions

A novel dimension reduction algorithm for the purpose of discrimination called kernel local linear discriminate (KLLD) has been proposed in this paper. The most prominent property of KLLD is the complete preservation of both discriminate and local geometrical structures in the data. However, traditional dimension reduction algorithm can not properly preserve the discriminate structure. We have applied our algorithm to iris databases dimension reduction. The experiment demonstrated that our algorithm can extract the different kinds of iris features and is suitable for classification. And then we applied KLLD to machinery fault diagnosis. At first, the original feature space of rolling bear dataset was constructed by wavelet energy. Secondly, KLLD algorithm and other dimensionality reduction methods were used, respectively, in the original feature space. Finally, SVM was used for classification. The experiment shows that our method has excellent capability of clustering and dimension reduction.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by the National Nature Science Foundation of China (no. 51175316), the Specialized Research Fund for the Doctoral Program of Higher Education (no. 20103108110006), and Shanghai Science and Technology Commission Basic Research Project (no. 11JC1404100).

References

- [1] W. Y. Liu and J. G. Han, "Rolling element bearing fault recognition approach based on fuzzy clustering bispectrum estimation," *Shock and Vibration*, vol. 20, pp. 213–225, 2013.

- [2] G. F. Wang, X. L. Feng, and C. Liu, "Bearing fault classification based on conditional random field," *Shock and Vibration*, vol. 20, pp. 591–600, 2013.
- [3] S. Mika, G. Rätsch, J. Weston, B. Scholkopf, and K.-R. Müller, "Fisher discriminant analysis with kernels," in *Proceedings of the 9th IEEE Workshop on Neural Networks for Signal Processing (NNSP'99)*, pp. 41–48, August 1999.
- [4] Q.-S. Jiang, M.-P. Jia, J.-Z. Hu, and F.-Y. Xu, "Method of fault pattern recognition based on laplacian eigenmaps," *Journal of System Simulation*, vol. 20, no. 20, pp. 5710–5713, 2008.
- [5] W. Yang, C. Sun, and L. Zhang, "A multi-manifold discriminant analysis method for image feature extraction," *Pattern Recognition*, vol. 44, no. 8, pp. 1649–1657, 2011.
- [6] P. K. Yan, W. X. Zhang, and B. Turkbey, "Global structure constrained local shape prior estimation for medical image segmentation," *Computer Vision and Image Understanding*, vol. 117, pp. 1017–1026, 2013.
- [7] V. D. Silva and J. B. Tenenbaum, "Global versus local methods in nonlinear dimensionality reduction," *Advances in Neural Information Processing Systems*, vol. 15, pp. 705–712, 2003.
- [8] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [9] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [10] D. Cai, X. F. He, and K. Zhou, "Locally sensitive discriminant analysis," in *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI'07)*, V. Manuela, Ed., pp. 708–713, 2007.
- [11] F. Li, B. P. Tang, and R. S. Yang, "Rotating machine fault diagnosis using dimension reduction with linear local tangent space alignment," *Measurement*, vol. 46, pp. 2525–2539, 2013.
- [12] W. Zhang, W. Zhou, and B. Li, "Fault diagnosis approach based on fractal dimension LLE and Fisher discriminant," *Chinese Journal of Scientific Instrument*, vol. 31, no. 2, pp. 325–333, 2010.
- [13] Z. Li, X. Yan, C. Yuan, J. Zhao, and Z. Peng, "A new method of nonlinear feature extraction for multi-fault diagnosis of rotor systems," *Noise & Vibration Worldwide*, vol. 41, no. 10, pp. 29–37, 2010.
- [14] B. Li and Y. Zhang, "Supervised locally linear embedding projection (SLLEP) for machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 25, no. 8, pp. 3125–3134, 2011.
- [15] Y. Lei, Z. He, and Y. Zi, "A new approach to intelligent fault diagnosis of rotating machinery," *Expert Systems with Applications*, vol. 35, no. 4, pp. 1593–1600, 2008.
- [16] Y. W. Pang, L. Zhang, and Z. K. Liu, "Neighborhood preserving projections (NPP): a novel linear linear dimension reduction method," in *Advances in Intelligent Computing*, vol. 3644 of *Lecture Notes in Computer Science*, pp. 117–125, Springer, Berlin, Germany, 2005.
- [17] H. F. Li, T. Jiang, and K. Zhang, "Efficient and robust feature extraction by maximum margin criterion," *IEEE Transactions on Neural Networks*, vol. 17, no. 1, pp. 157–165, 2006.
- [18] S. L. Jiang, J. Y. Zhang, and J. Xu, "Discriminative analysis algorithm for linear local tangent spaces," *Journal of Huazhong University of Science and Technology*, vol. 38, no. 10, pp. 66–69, 2010.
- [19] G. H. Wen, L. J. Jiang, and J. Wen, "Using locally estimated geodesic distances to improve Hessian local linear embedding," *CAAI Transactions on Intelligent Systems*, vol. 3, pp. 429–435, 2008.
- [20] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, article 27, 2011.
- [21] O. Kouropteva, O. Okun, and A. Hadid, "Beyond locally linear embedding algorithm," MVG-01-2002, Machine Vision Group, University of Oulu, 49, 2002.