Research Article

Bearing Condition Recognition and Degradation Assessment under Varying Running Conditions Using NPE and SOM

Shaohui Zhang¹ and Weihua Li^{1,2}

¹ School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510640, China

² State Key Laboratory for Manufacturing System Engineering, Xi'an Jiaotong University, Xi'an 710049, China

Correspondence should be addressed to Weihua Li; whlee@scut.edu.cn

Received 7 March 2014; Accepted 3 April 2014; Published 5 May 2014

Academic Editor: Ruqiang Yan

Copyright © 2014 S. Zhang and W. Li. 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.

Manifold learning methods have been widely used in machine condition monitoring and fault diagnosis. However, the results reported in these studies focus on the machine faults under stable loading and rotational speeds, which cannot interpret the practical machine running. Rotating machine is always running under variable speeds and loading, which makes the vibration signal more complicated. To address such concern, the NPE (neighborhood preserving embedding) is applied for bearing fault classification. Compared with other algorithms (PCA, LPP, LDA, and ISOP), the NPE performs well in feature extraction. Since the traditional time domain signal denoising is time consuming and memory consuming, we denoise the signal features directly in feature space. Furthermore, NPE and SOM (self-organizing map) are combined to assess the bearing degradation performance. Simulation and experiment results validate the effectiveness of the proposed method.

1. Introduction

Bearings are among the most essential components in rotating machinery. They are frequently operated under high loading and severe conditions. And the defects often occur gradually on the bearing. In order to prevent unexpected bearing failures, many fault diagnosis methods have been explored for faults detection, such as temperature monitoring, oil analysis, and vibration analysis. Among them, vibration signal-based method has been extensively used for bearing condition monitoring. The traditional vibration signal processing methods include time domain analysis [1], frequency domain analysis (such as FFT transform and envelope demodulation [2]), and time-frequency domain analysis (such as wavelet transform [3] and WV distribution [4]). Many features can be extracted from vibration data using these methods. Feature extraction is implemented to serve as a preprocessor for the fault diagnosis and the performance assessment. The classical feature extraction methods, such as PCA (principal component analysis) [5] and LDA (linear discriminant analysis) [6], have been successfully applied in fault diagnosis.

However, PCA and LDA are preferable in applications where the data space is linear. As opposed to these techniques, some dimensionality reduction techniques were proposed for processing nonlinear data structure in recent years, such as NPE (neighbor preserving embedding) [7], LPP (locality preserving projections) [8], and ISOP (isometric projection) [9]. And these methods have been effectively used in machine condition monitoring and fault diagnosis. Yu [10] proposed a local and nonlocal preserving projection (LNPP) algorithm for machine fault diagnosis and performance prognostics. Yang et al. [11] adopted the principal manifold learning to reduce the noise of nonlinear time series. Jiang et al. [12] developed supervised Laplacian eigenmaps technique for gearbox fault classification. Li and Zhang [13] used supervised locally linear embedding projection for bearing fault diagnosis.

One of the primary difficulties in bearing condition monitoring is how to eliminate the noise influence. Generally, vibration signals are sampled as longtime series in order to improve the frequency resolution. The traditional denoising methods, like SVD (singular value decomposition) [14] and wavelet [15], are time consuming and memory consuming.

When various features are extracted from vibration data, the noise contained in the data is also transferred to the features. Therefore, we think that denoising these features directly can speed up the computation and also save the memory space.

In addition, studies of bearing fault diagnosis focused on the bearing faults under stable loading and rotational speeds, which cannot interpret the machine running conditions. The bearings were always running under variable loading and speeds, and the actual vibration signal is more complicated. How to process these complicated vibration signals is another key issue to perform bearing health assessment.

The objective of this work is to explore the effectiveness of unsupervised NPE algorithm on bearing fault diagnosis and degradation assessment. The experimental results illustrate that the proposed scheme is capable of identifying different fault modes and evaluating degradation performance.

The rest of the paper is organized as follows. Section 2 presents the basic theory of NPE algorithm briefly. In Section 3, the feature-denoising algorithm is presented and the NPE approach is used to recognize the bearing vibration signals measured under variable loading and speeds. In Section 4, NPE is combined with SOM (self-organizing map) to describe the bearing defect propagation. The conclusion in Section 5 closes the paper.

2. Neighborhood Preserving Embedding (NPE)

Locally linear embedding (LLE) does not require an iterative algorithm and just a few parameters need to be set, which leads to its application in fault diagnosis. However, the performance of the LLE is sensitive to the selection of the nearest neighbors. The NPE can avoid this disadvantage, since it is less sensitive to outliers than LLE. NPE aims at preserving the local manifold structure after dimension reduction, and it is a linear approximation of the LLE.

Given a set of N samples assembled in a matrix $\mathbf{X} = [x_1, x_2, x_3, \dots, x_N]$, size $M \times N$, a transformation matrix $\mathbf{A}$ can be constructed to project these N samples to images assembled in a matrix $\mathbf{Y} = [y_1, y_2, y_3, \dots, y_N]$, size $d \times N$ ($d \ll M$), where the r th column vector of $\mathbf{Y}$ corresponds to that of $\mathbf{X}$, respectively.

(1) Constructing an adjacency graph: calculate Euclidean distance between samples x_i and x_j , and use k -nearest neighbor to construct the adjacency graph $\mathbf{G}$. The distance $d(x_i, x_j)$ represents the edge connecting x_i and x_j , as

$$d(x_i, x_j) = \|x_i - x_j\|. \quad (1)$$

(2) Computing the weights: for each sample, it can be represented as a linear combination of the neighbors, and the weight matrix reflects the coefficients. In this step, we can compute the corresponding weight w_{ij} by minimizing the following weighted cost function $\varepsilon(\mathbf{w})$:

$$\varepsilon(\mathbf{w}) = \min \left\| x_i - \sum_{j=1}^k w_{ij} x_i^j \right\|^2, \quad (2)$$

where ε is the reconstruction error, w_{ij} is the weight of the j th neighbor of data x_i , with the constraint $\sum_{j=1}^k w_{ij} = 1$, and

x_i^j is the j th neighbor of data x_i . Let $(w_{i1}, w_{i2}, \dots, w_{ik})$ be a weight vector and obtain N -D vector $\mathbf{W}_i$ by adding zero for the nonneighbors. All these vectors are used to construct a matrix $\mathbf{W}$, size $N \times N$. Equation (2) can be rewritten as

$$\varepsilon(\mathbf{W}_i) = \min \|x_i - \mathbf{XW}_i^T\|^2. \quad (3)$$

Therefore,

$$\begin{aligned} \varepsilon(\mathbf{W}) &= \min \left\{ \sum_{i=1}^N \|x_i - \mathbf{XW}_i^T\|^2 \right\} \\ &= \min \{ \text{tr}(\mathbf{XMX}^T) \}, \end{aligned} \quad (4)$$

where tr is the trace of $\mathbf{XMX}^T$, $\mathbf{M} = (\mathbf{I} - \mathbf{W}^T)^T(\mathbf{I} - \mathbf{W}^T)$, and $\mathbf{I}$ is an $N \times N$ identity matrix.

Suppose the transformation is $\mathbf{Y} = \mathbf{A}^T \mathbf{X}$. For the purpose of removing an arbitrary scaling factor in the projection, we impose a constraint function as follows:

$$\mathbf{Y}\mathbf{Y}^T = 1 \implies \mathbf{A}^T \mathbf{X}\mathbf{X}^T \mathbf{A} = 1. \quad (5)$$

(3) Computing the projections and combining (4) and (5) yield the following generalized eigenvector problem:

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

$a_0, \dots, a_{d-1}$ are arranged according to their corresponding eigenvalues $\lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{d-1}$. Therefore, the embedding is performed as follows:

$$x_i \longrightarrow y_i = \mathbf{A}^T x_i, \quad (7)$$

where y_i is a d dimensional vector and $\mathbf{A} = (a_0, a_1, \dots, a_{d-1})$, size $M \times d$.

It should be pointed out that “labeled” data is often used for supervised learning while “unlabeled” data for unsupervised learning, and NPE can adopt both of them for learning process. The class information can be utilized to get a better dimensionality reduction. But in practical applications, it is often difficult to collect labeled instances. However, unlabeled data may be relatively easy to acquire. The focus of this paper is to demonstrate the effectiveness of NPE in bearing fault classification and degradation assessment, without class information.

3. Simulation and Experiments

3.1. Setup. In this section, the NPE is used to process the simulation signals and the bearing vibration data. Then, the INN (1-nearest neighbor) method is adopted to classify different bearing fault, because it is the simplest classification algorithm which can be used even with few samples, and it works very well in low dimensions for complex decision surfaces. The procedure of machine condition recognition and health assessment is illustrated in Figure 1.

In order to highlight the effectiveness of the NPE algorithm, the results of NPE are compared with those of unsupervised learning (LPP and PCA) and supervised learning (LDA, ISOP, and supervised NPE as SNPE), respectively.

FIGURE 1: The process of machine condition recognition and health assessment.

Case 1 (simulation). To describe the waveform generated by a rolling element bearing under the constant radial load with a single localized defect, the vibration signature can be expressed as (8) in [16]:

$$y(t) = y_d(t) y_q(t) y_r(t) y_e(t) + y_n(t), \quad (8)$$

where y_d is a series of impulses at bearing fault frequency, y_q is the bearing radial load distribution, y_r represents the bearing induced resonant frequency, y_e reflects the exponential decay due to damping, and y_n is the noise.

The rotational speed was set at 1100 rpm, and the vibration signals were measured at a constant sampling rate 120 kHz; the duration of each vibration signal was 42 seconds. The fault frequency was placed at 133 Hz and impact amplitudes were set as 0, 1, 1.5, and 2, which represented the normal, slight, moderate, and severe fault, respectively. White Gaussian noise with SNR = -2 dB was added to the signal.

Case 2 (bearing conditions recognition under varying working conditions). The data used in this work were obtained from the Case Western Reserve University Bearing Data Centre [17]. All the experiments were repeated for different loading conditions: 1, 2, and 3 hp at rotational speeds ranging from 1730 rpm to 1772 rpm. The data were sampled at a rate of 48 kHz and the duration of each vibration signal was 10 seconds. Defects were introduced into the rolling element bearing at drive-end by using electrodischarge machining in the following configurations:

- (i) healthy bearing being used as a baseline,
- (ii) inner race defect (0.007 inch, 0.014 inch, 0.021 inch in diameter, and 0.011 inch in depth),
- (iii) outer race defect (0.007 inch, 0.014 inch, 0.021 inch in diameter, and 0.011 inch in depth),
- (iv) ball defect (0.007 inch, 0.014 inch, 0.021 inch in diameter, and 0.011 inch in depth).

So there are totally 10 kinds of bearing conditions, and each defect has three different levels under varying loading and speeds, and there are 600 vibration data sampled in total (60 datasets per condition). Then 20 features (including time domain and frequency domain features, as in Table 1) were extracted from the data. Therefore, the original feature data is of size (600×20) in high-dimension space.

TABLE 1: Bearing feature generation.

Time domain features	Frequency domain features
Mean square	A_{f_n} amplitude at rotating frequency f_n
Kurtosis	A_{2f_n} amplitude at $2f_n$
Mean	A_{BPFI} amplitude at BPFI
Skewness	A_{2BPFI} amplitude at 2BPFI
Peak value	A_{BPFO} amplitude at BPFO
Rms	A_{2BPFO} amplitude at 2BPFO
Waveform factor	A_{FTF} amplitude at FTF
Crest factor	A_{2FTF} amplitude at 2FTF
Impulse factor	A_{BSF} amplitude at BSF
Clearance factor	A_{2BSF} amplitude at 2BSF

Case 3 (bearing degradation assessment using NPE and SOM). The data under normal state and ball defects (0.007 inch, 0.014 inch, 0.021 inch in diameter, and 0.011 inch in depth) were selected to analyze the bearing degradation trend.

3.2. Simulation Results Analysis. The simulation signal under different impacts is shown in Figure 2. It can be seen that there is little difference between the normal signal and slight fault one, even the moderate fault, due to the impact of noise. As for the severe fault bearing signal, the impact amplitude is larger than those of others. Each vibration signal was divided into forty segments and the corresponding features as listed in Table 1 were extracted, and there were 160 20-D feature data sets obtained in total. The feature curves before and after denoising are shown in Figure 3. It can be observed that denoising features in the feature space can reduce the margin fluctuation of the curves, which is beneficial to the fault classification, especially for features as mean, skewness, impulse factor, clearance factor, and most of frequency features.

For verifying the generalization performance of the proposed model, 2 random subsets with 25% and 50% of the simulation data were used for training, and the rest for testing. This was aimed at examining the performance of the NPE dealing with data sets for which it was not sufficiently trained. The training dataset was used to obtain the transformation matrix A . Then the testing data were input into the learning machine to construct the feature subspace and used the INN to calculate misclassification rate. This process was repeated for 20 times, and the result was the

FIGURE 2: Simulation of vibration signals under different impacts.

FIGURE 3: Features variation before and after denoising (BD: before denoising and AD: after denoising).

FIGURE 4: Simulation result: the error rate of different algorithms without denoising.

FIGURE 5: Simulation result: the error rate of different algorithms with denoising.

average of 20 processes. The error rate versus dimension is shown in Figure 4.

As can be seen from Figure 4(a), when 25% samples were used for training, the misclassification of unsupervised NPE is lower than that of others. In 2-dimension space, the LPP, LDA, and ISOP perform better, and their error rates

are 3.912%, 4.958%, and 3.917%, respectively. In Figure 4(b), given 50% samples for training, both of algorithms can achieve good results, and the error rates of them are less than 10%. The aforementioned learning methods perform better in 2D space, and the misclassification rate of SNPE is the maximum, 9.875%. However, the performance of NPE

is stable and acceptable; the error rate is ranging from 5.792% to 6.292% (25% training samples) or 2.688%–3.125% (50% training samples). When the dimension increases, it is superior to other algorithms.

To speed up computation and save memory space, the feature samples are denoised directly using singular value decomposition (SVD). The steps of SVD for noise reduction are as follows.

- (1) The vibration signal $\mathbf{X}$ can be transformed into an $n \times m$ matrix $\mathbf{Q}$ by using phase space reconstruction with the time delay of “one unit.” The SVD of $\mathbf{Q}$ is a factorization of the formula $\mathbf{Q} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where $\mathbf{U}$ is an $n \times n$ real or complex unitary matrix, $\mathbf{V}^T$ (the conjugate transpose of $\mathbf{V}$) is an $m \times m$ real or complex unitary matrix, and $\mathbf{\Sigma}$ is a diagonal matrix of size $n \times m$, and its diagonal values are the singular values of matrix $\mathbf{Q}$. The singular value reflects the energy concentration between signals and noises. The greater the value is, the less influence the noise affects. Keep great values and set all the small values as zero, which will effectively improve signal-to-noise ratio and obtain a diagonal matrix $\mathbf{\Sigma}'$.
- (2) Calculate space matrix again as $\mathbf{Q}' = \mathbf{U}\mathbf{\Sigma}'\mathbf{V}^T$ and restore signal $\mathbf{X}'$ to one-dimensional series after noise reduction.

After feature denoising, the learning and classification were repeated using NPE and other algorithms. The error rates of different algorithms with denoising feature samples are shown in Figure 5.

Comparing Figure 4 with Figure 5, it can be observed that denoising feature sets can depress the misclassification effectively. Both of the algorithms can achieve good results, and all the error rates are less than 4%. As can be seen from Figure 5, the performance of NPE is stable and the misclassification of NPE is the lowest of these algorithms; its error rate is ranging from 0.0833% to 0.2917% (25% training samples) and 0%–0.0625% (50% training samples). The largest error rate is of SNPE, 3.958% with 25% training samples, while that is 2.063% of LDA with 50% training samples.

3.3. Experiment Results and Discussion. Bearing datasets in Case 2 were used to validate the proposed approach in bearing condition recognition.

For verifying the generalization performance of the proposed model, 2 random subsets with 30% and 50% of the data were used for training, and the rest for testing. This process was repeated for 20 times, and the result was the average of 20 processes. The error rate versus dimension is revealed in Figure 6.

As depicted in Figure 6, the misclassification rate decreased with the dimension increasing and the training samples growing, and the supervised methods are superior to the unsupervised ones. It is obvious that the misclassification of LDA, ISOP, and SNPE is lower than those of unsupervised methods. But in reality, it is often difficult or expensive to collect labeled instances. However, unlabeled data may be

relatively easy to acquire. The unsupervised learning methods are more appropriate for practical application.

The NPE achieves the 90% classification correctness (error rate 9.767%, 50% training samples) in the 4-dimension feature space. It can be seen that, for multifault classification, the dimension of feature space affects the classifying results. In 2D and 3D space, most of the methods cannot get the good classification, except for the SNPE (error rate 9.917%, 50% training samples in 3D space).

In order to validate the feature denoising, we denoised features directly by SVD, and the results are shown in Figure 7.

Comparing Figure 6 with Figure 7, it can be observed that denoising feature sets can decrease the misclassification rate effectively. Both of the algorithms can achieve better results than those before denoising, and all the error rates are less than 10% when the dimension is more than 2. As can be seen from Figure 7, the performance of NPE is also agreeable and acceptable. The misclassification of NPE is the lowest among unsupervised algorithms (PCA, LPP, and NPE) when the subspace dimension exceeds 3, and its error rate is ranging from 4.452% to 6.381% (30% training samples) and 3.45% to 5.88% (50% training samples). Whereas in 3D feature space, it achieves the best classification result (the error rate 6.381%) with only 30% training samples involved in learning.

4. Bearing Degradation Performance Assessment Using NPE and SOM

In this section, the bearing defect propagation was investigated by using SOM. Rolling element bearings with ball defect (normal, fault size with 0.007 inch, 0.014 inch and 0.021 inch in diameter, and 0.011 inch in depth) under different loadings (1 hp, 2 hp, and 3 hp) were used to implement the degradation assessment, 240 datasets in total. Firstly, the unsupervised NPE was used for feature extraction. Then the MQE (minimum quantization error) of the SOM can be used to observe the bearing degradation trend.

SOM (self-organizing map) was developed by Kohonen [18], and it is a general unsupervised tool for clustering. It consists of neurons located on a regular, usually 2D grid of map units. And for similar samples in the input space, they can be mapped to the neighbor neurons in the output space. SOM has been proven useful in gearbox condition monitoring [19], degradation assessment of rolling element bearing [20], and so on.

In fact, it is often difficult to collect datasets which can represent the whole failure space. Meanwhile, healthy samples can be obtained relatively easy, which can be used to characterize the normal state. And the deviation from the normal feature space can reflect degradation detection. A quantitative degradation index as the MQE of SOM can be obtained by depending on distance between the normal state and the current process, which can be normalized by converting the MQE into confidence value (CV) ranging from 0 to 1. At first, only the healthy data are needed for training SOM model. The new sample is input into the model and compared with the weight vectors of all map units,

FIGURE 6: Bearing fault classification: the error rate of different algorithms without denoising.

FIGURE 7: Bearing fault classification: the error rate of different algorithms with denoising.

and when the smallest difference exceeds a predetermined threshold, this sample is probably in a fault situation.

Quantization error shows how accurately the neurons of the trained network respond to the given input samples and is also the average distance between the data vector x_i and the BMU (best matching unit). The MQE could be calculated

according to (9), and more detail information about the algorithm can be found in [20], as follows:

$$\text{MQE}(i) = \|x_i - h_{\text{BMU}}\|, \quad (9)$$

FIGURE 8: Residual curve of bearing running state.

where h_{BMU} represent the BMU of healthy data, and x_i denotes the test samples.

To investigate the bearing degradation performance, we need to represent the bearing running tendency in a low dimensional space. Therefore, NPE was used to decrease the dimension of feature space. The intrinsic dimension of the original feature space describes how many variables are needed to represent the bearing running. There are two kinds of methods to estimate the intrinsic dimension. One is the eigenvalue method, such as PCA, which determines the intrinsic dimension through the number of eigenvalues greater than a given threshold; however, it failed on the nonlinear manifold. And the other is the geometric one, such as ISOMAP (isometric feature mapping), which exploits the intrinsic geometry based on the nearest neighbor distances. It provides residual error curves that can be “eyeballed” to estimate the intrinsic dimension [21], and residual variance decreases with added dimensions. The intrinsic dimension of the data can be estimated by finding the “elbow” at which the residual curve ceases to decrease significantly with dimension increases, as shown in Figure 8. It can be seen that the “elbow” point is at the place where the dimension is three. It means that the 3D subspace can be used to describe bearing running trend.

Therefore, three new features can be obtained by the NPE from the original feature sets. To eliminate the noise influence, the original features were denoised, and the variation of these features was described in Figure 9(b), in comparison with that before denoising (as shown in Figure 9(a)). It can be seen that, after feature denoising, the intraclass samples gathered together and the interclass ones were separated relatively.

A random subset with 50% of the data in Case 3 was selected for training, while the remaining for testing. At first, the NPE was trained to obtain the transformation matrix $\mathbf{A}$. Then the testing data were input into the learning machine to construct the feature subspace. To validate the MQE’s

capability of degradation detection, SOM was firstly trained by the selected normal datasets from subspace. Then the test data in subspace was input to the learnt SOM, and the confidence value (MQE) was calculated to measure the deviation from the normal state. This process was repeated for 20 times, and the result was the average of 20 processes.

We compared the MQE of the original feature space with those of NPE feature subspace. Figure 10 shows the MQE curves of the original features, the NPE features, and the NPE features with denoising, respectively.

Comparing Figure 10(a) with Figure 10(b), it can be observed that the bearing degradation tendency can be described in the NPE subspace, validating that the intrinsic dimension is determined correctly. However, it is still unable to distinguish the ball defect BD014 (0.014 inch) from defect BD021 (0.021 inch). The confidence value of BD021 (1 hp, 1772 rpm) data is lower than that of BD014 (1 hp, 2 hp, and 3 hp), due to the noise. Even the confidence values of the same degradation state BD014 were varying greatly under different working conditions.

Therefore, the original features were denoised in order to eliminate the noise effect on vibration features, where the result was shown in Figure 10(c). It can be seen that the confidence value varies with different state obviously, especially for the moderate defect BD014 and the severe defect BD021. The worse the rolling element degraded, the bigger the confidence value (MQE) increased. At the same time, the confidence value of the same state fluctuated narrowly, especially for BD014, while the mean value of BD021 (1 hp, 1772 rpm) was almost the same as that of BD014 (3 hp, 1730 rpm). The results indicated that noise reduction facilitates a reliable rolling bearing performance prediction, and the proposed NPE-MQE can effectively assess the bearing degradation performance.

5. Conclusions

To investigate the fault recognition under varying working conditions, the NPE was adopted to perform dimension reduction and INN was used to classify different faults. Furthermore, the NPE and SOM were combined together for bearing degradation performance evaluation. The work in this study can be summarized as follows.

- (1) Denoising features can speed up the computation, save memory space, and improve the recognition accuracy effectively, which is very helpful for fault classification.
- (2) Simulation and experiment results demonstrate that the NPE is capable of extracting discriminative features, even lacking training samples. It is beneficial to both fault classification and degradation assessment.
- (3) The proposed NPE-MQE method can be used to assess the bearing degradation performance, and the confidence value can depict the bearing degradation process efficaciously.

FIGURE 9: New features extracted by NPE.

FIGURE 10: The MQE variation in time.

Conflict of Interests

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

Acknowledgments

This work described in this paper is supported in part by the National Natural Science Foundation of China (Grant no. 51075150) and State Key Laboratory for Manufacturing System Engineering (SKLMS2014008).

References

- [1] F. Gao and Y. Lu, "A Kalman-filter based time-domain analysis for structural damage diagnosis with noisy signals," *Journal of Sound and Vibration*, vol. 297, no. 3–5, pp. 916–930, 2006.
- [2] H. Zoubek, S. Villwock, and M. Pacas, "Frequency response analysis for rolling-bearing damage diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 12, pp. 4270–4276, 2008.
- [3] J. Kim, D. E. Welcome, R. G. Dong, W. J. Song, and C. Hayden, "Time-frequency characterization of hand-transmitted, impulsive vibrations using analytic wavelet transform," *Journal of Sound and Vibration*, vol. 308, no. 1–2, pp. 98–111, 2007.
- [4] W. J. Staszewski, K. Worden, and G. R. Tomlinson, "Time-frequency analysis in gearbox fault detection using the Wigner-Ville distribution and pattern recognition," *Mechanical Systems and Signal Processing*, vol. 11, no. 5, pp. 673–692, 1997.
- [5] I. Jolliffe, *Principal Component Analysis*, Springer, New York, NY, USA, 1986.
- [6] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, New York, NY, USA, 1973.
- [7] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV '05)*, pp. 1208–1213, October 2005.
- [8] X. He and P. Niyogi, "Locality preserving projections," in *Advances in Neural Information Processing Systems*, vol. 16, pp. 153–160, MIT Press, Cambridge, Mass, USA, 2004.
- [9] D. Cai, X. He, and J. Han, "Isometric projection," in *Proceedings of the 22nd Conference on Artificial Intelligence (AAAI '07)*, pp. 528–533, July 2007.
- [10] J. Yu, "Local and nonlocal preserving projection for bearing defect classification and performance assessment," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 5, pp. 2363–2376, 2012.
- [11] J. Yang, J. Xu, D. Yang, and M. Li, "Noise reduction method for nonlinear time series based on principal manifold learning and its application to fault diagnosis," *Chinese Journal of Mechanical Engineering*, vol. 42, no. 8, pp. 154–158, 2006.
- [12] Q. Jiang, M. Jia, J. Hu, and F. Xu, "Machinery fault diagnosis using supervised manifold learning," *Mechanical Systems and Signal Processing*, vol. 23, no. 7, pp. 2301–2311, 2009.
- [13] 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.
- [14] K. Shin, J. K. Hammond, and P. R. White, "Iterative SVD method for noise reduction of low-dimensional chaotic time series," *Mechanical Systems and Signal Processing*, vol. 13, no. 1, pp. 115–124, 1999.
- [15] Y.-F. Sang, D. Wang, J.-C. Wu, Q.-P. Zhu, and L. Wang, "Entropy-based wavelet de-noising method for time series analysis," *Entropy*, vol. 11, no. 4, pp. 1123–1147, 2009.
- [16] Y.-F. Wang and P. J. Kootsookos, "Modeling of low shaft speed bearing faults for condition monitoring," *Mechanical Systems and Signal Processing*, vol. 12, no. 3, pp. 415–426, 1998.
- [17] K. A. Loparo, Bearing Data Center, Case Western Reserve University, 2013, <http://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website>.
- [18] T. Kohonen, *Self-Organizing Maps*, Springer, Berlin, Germany, 3rd edition, 2001.
- [19] G. Liao, S. Liu, T. Shi, and G. Zhang, "Gearbox condition monitoring using self-organizing feature maps," *Proceedings of the Institution of Mechanical Engineers C: Journal of Mechanical Engineering Science*, vol. 218, no. 1, pp. 119–130, 2004.
- [20] H. Qiu, J. Lee, J. Lin, and G. Yu, "Robust performance degradation assessment methods for enhanced rolling element bearing prognostics," *Advanced Engineering Informatics*, vol. 17, no. 3–4, pp. 127–140, 2003.
- [21] E. Levina and P. J. Bickel, "Maximum likelihood estimation of intrinsic dimension," in *Advances in Neural Information Processing Systems*, vol. 17, MIT Press, Cambridge, Mass, USA, 2004.