Available online at www.ijpe-online.com

International Journal of Performability Engineering

vol. 15, no. 10, October 2019, pp. 2657-2666
DOI: 10.23940/ijpe.19.10.p11.26572666

Local and Global SR for Bearing Sensor-based Vibration Signal Classification

Shaohui Zhang[a], Man Wang[a], Canyi Du[b,*], and Edgar Estupinan[c]

[a]*School of Mechanical Engineering, Dongguan University of Technology, Dongguan, 523808, China*
[b]*School of Automobile and Transportation Engineering, Guangdong Polytechnic Normal University, Guangzhou, 510665, China*
[c]*Department of Mechanical Engineering, University of Tarapacá, Arica, Chile*

Abstract

Spectral regression (SR) is a method of feature extraction that realizes dimension reduction by the least squares method and can avoid eigen-decomposition of dense matrices. However, it only considers the affinity graph and misses the global information. In this paper, a novel feature extraction algorithm, called local and global spectral regression (LGSR), is proposed and applied to extract fault features from frequency-domain and time-domain features of vibration signals of bearing sensors. LGSR, which is the development of SR, is able to discover both local and global information of data manifold. Compared with other similar approaches (such as NPE, PCA, and SR), experiments of bearing defect classification validate that LGSR shows better ability to extract identity information for machine defect classification.

Keywords: spectral regression; sensor-based signals; feature extraction; conditions classification

(Submitted on June 15, 2019; Revised on July 12, 2019; Accepted on August 11, 2019)

1. Introduction

The bearing is one of the most essential components in rotating machinery. During operation, bearings are often subjected to high loading and severe conditions, which lead to the occurrence of bearing defects. Therefore, in order to prevent accidental failures of bearings, researchers have developed many fault diagnosis methods to detect and identify faults. Among them, the method based on vibration signal analysis is most suitable and effective, and it has been widely used in various machine fault diagnosis.

The analysis methods based on vibration signals are 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]). Based on these methods, many features can be extracted, and the defect information is contained in the features. However, the original feature set has high dimensionality, and it is critical to extract the fault information hiding in these features.

PCA (principal component analysis) [5] is one of the feature extraction algorithms, and it has been applied in bearing fault diagnosis. LDA (linear discriminant analysis) [6] is another feature extraction technique, but it is not appropriate for fault diagnosis since it needs sample labels. Furthermore, both PCA and LDA are linear projection techniques, which cannot provide a good mapping function from high dimensional spaces to low dimensional spaces for training and testing data simultaneously.

Many mechanical systems are often characterized by nonlinear behaviors, due to instantaneous variations in friction, damping, and load [7]. This prompted us to consider reducing the number of dimensions based on multiple methods. Some manifold learning algorithms have been proposed for dimension reduction and have been effectively used to monitor machine running and diagnose machine faults. To obtain more information, Yu [8] proposed a novel feature extraction

* Corresponding author.
E-mail address: 229457893@qq.com

method, named LNPP, for bearing fault diagnosis. Wang et al. [9] proposed a waveform feature manifold (WFM) method to extract the weak signature for bearing fault diagnosis. Ding [10] proposed improved fast TFM (FTFM) for rotating machinery fault diagnosis. To enhance the transient features, Ding et al. [11] proposed the time-frequency manifold image synthesis method for bearing and gear fault diagnosis.

In recent years, a feature extraction technology called SR (spectral regression), which is fundamentally based on regression and spectral graph analysis [12], was proposed. First, an affinity graph over points is constructed to discover the intrinsic discriminant structure in the data, and then the ordinary regression is applied to learn the embedding function. SR has been used in a variety of applications, and Xia [13] used SR for bearing fault diagnosis.

Since SR only considers the affinity graph, which may lead to missing some useful knowledge, we consider the local and global structure of the data manifold. Therefore, in this paper, a novel technology, named local and global spectral regression (LGSR), is proposed for feature extraction of bearing faults. First, the affinity and global map are constructed on the point to discover the intrinsic discriminant structure in the data, and then ordinary regression is applied to learn the embedded function.

In addition, the study of bearing fault diagnosis mainly focuses on bearing failures under stable loads and speeds but does not explain the actual operation of the machine. In fact, bearings always operate at different loads and speeds, and sensor-based vibration signals are complex. Therefore, this paper will apply LGSR, SR, LPP (locality preserving projections), PCA, and NPE (neighborhood preserving embedding) to fault feature extraction and analyze the applicability of these methods.

The rest of this paper is organized as follows: Section 2 introduces the basic theory of the LGSR algorithm. The experimental and analytical description and evaluation of the effectiveness of the proposed method using bearing sensor vibration signals are given in Section 3. Section 4 concludes the paper.

2. Local and Global Spectral Regression (LGSR)

In this section, local and global spectral regression (LGSR), which is the development of SR, is presented. The algorithmic procedure is as follows:

Given a set of N points combined in a matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N]$ in R_m, find a transformation matrix $\boldsymbol{A}$ that maps these N points to $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_N]$ in R_d ($d << m$), where the k^{th} column vector of $\boldsymbol{Y}$ corresponds to the k^{th} column vector of $\boldsymbol{X}$.

(1) Constructing the distance. Calculating the Euclidean distance between samples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. The edge connecting $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ denotes the distance $d(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

$$d(\boldsymbol{x}_i, \boldsymbol{x}_j) = \| \boldsymbol{x}_i - \boldsymbol{x}_j \| \tag{1}$$

(2) Choosing the weights. We define the affinity matrix $\boldsymbol{W}$ and global matrix $\boldsymbol{S}$ as follows. $\boldsymbol{W}$ and $\boldsymbol{S}$ are a sparse symmetric $N \times N$ matrix.

a) For affinity matrix $\boldsymbol{W}$, we can weight the edge as

$$W_{ij} = \begin{cases} \exp(-\left\| \boldsymbol{x}_i - \boldsymbol{x}_j \right\|^2 / 2\sigma^2), & \text{if } \boldsymbol{x}_i (\text{or } \boldsymbol{x}_j) \text{ is among } k_l \\ & \text{nearest neighbor of } \boldsymbol{x}_j (\text{or } \boldsymbol{x}_i) \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

Where $\sigma \in R$.

b) For global matrix $\boldsymbol{S}$, we set the weight simply as

$$S_{ij} = \begin{cases} -1, & \text{if } \boldsymbol{x}_i (\text{or } \boldsymbol{x}_j) \text{ is among } k_g \\ & \text{farthest distance of } \boldsymbol{x}_j (\text{or } \boldsymbol{x}_i) \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

(3) Computing the projections

A reasonable criterion for preserving the local information is to minimize the follow cost function $\varepsilon(\boldsymbol{y})$.

$$\begin{aligned}\varepsilon(\boldsymbol{y}) &= \min \frac{1}{2}\sum_{ij}(\boldsymbol{y}_i - \boldsymbol{y}_j)^2 W_{ij} \\ &= \min[\boldsymbol{y}^{\mathrm{T}}(\boldsymbol{D}^l - \boldsymbol{W})\boldsymbol{y}] \\ &= \min(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{L}^l\boldsymbol{y})\end{aligned} \tag{4}$$

Where $\boldsymbol{L}^l = \boldsymbol{D}^l - \boldsymbol{W}$ and $\boldsymbol{D}^l$ is a diagonal matrix and $D_{ii}^l = \sum_j W_{ji}$.

In order to extract the global information, we hope that the farthest distance points maintain their distance. The criterion for choosing a "good" map is to maximize the following cost function $\phi(\boldsymbol{y})$:

$$\begin{aligned}\phi(\boldsymbol{y}) &= \max \frac{1}{2}\sum_{ij}(\boldsymbol{y}_i - \boldsymbol{y}_j)^2 S_{ij} \\ &= \max[\boldsymbol{y}^{\mathrm{T}}(\boldsymbol{D}^g - \boldsymbol{S})\boldsymbol{y}] \\ &= \max(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{L}^g\boldsymbol{y})\end{aligned} \tag{5}$$

Where $\boldsymbol{L}^g = \boldsymbol{D}^g - \boldsymbol{S}$, $\boldsymbol{D}^g$ is a diagonal matrix, and $D_{ii}^g = \sum_j S_{ji}$.

For the purpose of clustering, a good projection should be one that minimizes $\varepsilon(\boldsymbol{y})$ and maximizes $\phi(\boldsymbol{y})$ simultaneously [8]. Thus, we find the good projection matrix $\boldsymbol{y}$ by combining Equations (4) and (5).

$$\boldsymbol{y}^* = \arg\min \frac{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{L}^l\boldsymbol{y}}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{L}^g\boldsymbol{y}} \tag{6}$$

The optimal $\boldsymbol{y}$ can be obtained by solving the minimum eigenvalue eigen-problem.

$$\boldsymbol{L}^l\boldsymbol{y} = \lambda\boldsymbol{L}^g\boldsymbol{y} \tag{7}$$

The eigen-problem in Equation (7) is solved to obtain $\boldsymbol{y}$ and the transform matrix $\boldsymbol{A}$, where $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_d]$ is the eigenvectors that are sorted according to their eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$. Once $\boldsymbol{A}$ is obtained, the linear features can be extracted by

$$\boldsymbol{y} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} \tag{8}$$

In some cases, this eigen-problem formulation can provide a more numerically stable solution, such as in [8]. In the rest of this section, we will develop an SR algorithm based on Equation (8).

(4) Spectral regression

In reality, there may not exist a vector $\boldsymbol{a}$ that satisfies $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X} = \boldsymbol{y}$. One possible approach is to find the a that best satisfies the equation in the sense of least squares.

$$\boldsymbol{a}_k = \arg\min_{\boldsymbol{a}} \sum_{i=1}^{N}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_i - \boldsymbol{y}_i)^2 \tag{9}$$

In the situation that the number of samples is smaller than the number of features, Equation (9) is not suitable. The most

popular way to solve this problem is to impose a penalty on the norm of $\boldsymbol{a}$.

$$\boldsymbol{a}_k = \arg\min_{\boldsymbol{a}}[\sum_{i=1}^{N}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_i - \boldsymbol{y}_i)^2 + \alpha \| \boldsymbol{a} \|^2] \tag{10}$$

$\alpha \geq 0$ is a parameter that controls the amount of contraction.

(5) Embedding

In order to obtain $\boldsymbol{a}_k$, the following Gaussian elimination method can be used to solve the following linear equations:

$$(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} + \alpha \boldsymbol{I})\boldsymbol{a}_k = \boldsymbol{X}\boldsymbol{y}_k \tag{11}$$

Where $\boldsymbol{I}$ is an $N \times N$ identity matrix.

Let $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_d]$ be the transformation matrix with size $N \times d$. The samples can be embedded into d dimensional subspace by

$$\boldsymbol{x} \to \boldsymbol{z} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} \tag{12}$$

3. Bearing Condition Recognition based on LGSR

3.1. Original Feature Space Generation

In this paper, multi-domain features are adopted. Specifically, time domain statistical features, including *mean square* (f_1), *kurtosis* (f_2), *mean* (f_3), *skewness* (f_4), *peak value* (f_5), *RMS* (f_6), *waveform factor* (f_7), *crest factor* (f_8), *impulse facto r*(f_9), and *clearance factor* (f_{10}) [14], are adopted for machine fault diagnosis. For bearing, the characteristic features related to the characteristic frequencies of bearing components can be obtained according to their parameters, such as A_{fn} (f_{11}, *amplitude at rotating frequency*), A_{2fn} (f_{12}, *amplitude at double rotating frequency*), A_{BPFI} (f_{13}, *amplitude of ball-pass frequency of inner race*), A_{2BPFI} (f_{14}, *amplitude of double ball-pass frequency of inner race*), A_{BPFO} (f_{15}, *amplitude of ball-pass frequency of outer race*), A_{2BPFO} (f_{16}, *amplitude of double the ball-pass frequency of outer race*), A_{FTF} (f_{17}, *amplitude of fundamental train frequency*), A_{2FTF} (f_{18}, *amplitude of double fundamental train frequency*), A_{BSF} (f_{19}, *amplitude of ball-spin frequency*), and A_{2BSF} (f_{20}, *amplitude of double ball-spin frequency*) [15-16]. Because the component at the defect frequency in the FFT spectrum is often drowned out by other strong vibration components, especially at the initial stage of defect, the envelope spectrum is used to extract the frequency characteristics. A total of 20 original features are shown in Table 1 and used as input to the dimensionality reduction model. The techniques presented in this article are based on 20 original features, but they are not limited to them.

Table 1. Bearing feature generation

Time domain features	**Frequency domain features**
f_1 Mean square	f_{11} Amplitude at f_n
f_2 Kurtosis	f_{12} Amplitude at 2 f_n
f_3 Mean	f_{13} Amplitude at BPFI
f_4 Skewness	f_{14} Amplitude at 2BPFI
f_5 Peak value	f_{15} Amplitude at BPFO
f_6 RMS	f_{16} Amplitude at 2 BPFO
f_7 Waveform factor	f_{17} Amplitude at FTF
f_8 Crest factor	f_{18} Amplitude at 2 FTF
f_9 Impulse factor	f_{19} Amplitude at BSF
f_{10} Clearance factor	f_{20} Amplitude at 2 BSF

3.2. Data Generation

Case I. The vibration data were obtained from the Case Western Reserve University Bearing Data Centre [17]. As shown in Figure 1, the test stand consists of a motor (left), a torque transducer/encoder (center), a dynamometer (right), and control electronics (not shown). The vibration signal was collected using accelerometer sensors and a 16-channel DAT recorder.

The test bearings (the parameters are shown in Table 2) support the motor shaft. All experiments were repeated under different load conditions: 1, 2, and 3 hp, with speeds ranging from 1730rpm to 1772rpm. The sampling frequency was 48 kHz, and the duration of each accelerometer sensor vibration signal was 10 seconds. In the following configuration, the defect of the rolling element bearing is generated at the driving end by electric discharge machining:

- Healthy bearing
- Inner race defect (0.007 inch, 0.014 inch, 0.021 inch in diameter, 0.011 inch in depth)
- Outer race defect (0.007 inch, 0.014 inch, 0.021 inch in diameter, 0.011 inch in depth)
- Ball defect (0.007 inch, 0.014 inch, 0.021 inch in diameter, 0.011 inch in depth)

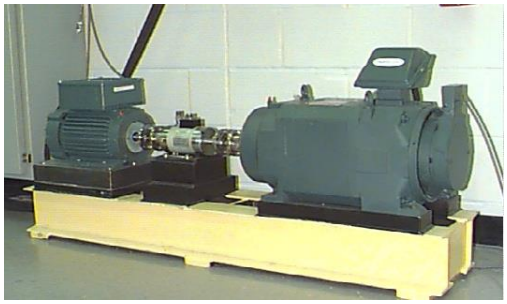


Figure 1. The test rig

Table 2. Bearing parameters

Parameters	**Values**	**Passing frequency** (Hz, multiples of rotating frequency *fn*)
Inner race diameter	25 mm	BPFO = 5.4152*fn*
Outer race diameter	52mm	BPFI = 2.0472*fn*
Width	15mm	BSF = 4.7135*fn*
Number of rollers	9	
Ball diameter	7.94mm	FTF = 0.39828*fn*
Pitch diameter	39.11mm	

Therefore, there were 10 bearing conditions, each having three different levels at variable loads and speeds, and a total of 600 vibration data samples (60 data sets per condition). Then, 20 features (including time domain and frequency domain features, as shown in Table 1) were extracted from the data samples. Therefore, the original feature data had a size of 600 × 20 in high-dimensional space.

Case II. As shown in Figure 2, the test rig consisted of the experiment platform, roller bearing (N205M) with inner and outer ring defect, BBM SCADA multi-analyzer system (PCB accelerometer sensors and a data acquisition system), and a thinkPad computer. The accelerometer was placed at the top of the bearing housing, and the rotation speed was kept constant at 500, 800, and 1100 RPM separately by an AC motor coupled to the shaft via rub belts. The sampling frequency was 12 kHz. The defect of the bearing was simulated by linear cutting with the following conditions:

- Healthy bearing
- Inner race defect (1mm in diameter, 1mm in depth)
- Outer race defect (1mm in diameter, 1mm in depth)

There was a total of 360 vibration data samples (40 data sets per condition), and 20 features could be extracted from the data samples. Therefore, the original feature data had a size of 360 × 20 in high-dimension space.

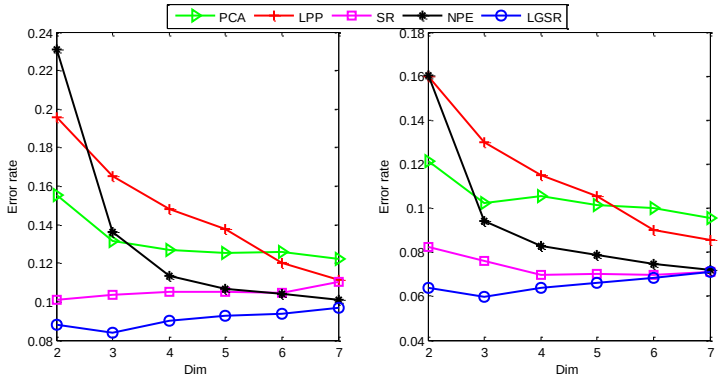
4. Experiment Result Analysis

4.1. Result Analysis of Case I

To verify the generalization performance of the proposed method, two random subsets with 30%, 50% samples were trained and the remaining samples were used for testing. The purpose of this was to examine the performance of LGSR in dealing with its undertrained data sets. The transform matrix $\boldsymbol{A}$ was obtained by the training data set, the testing data was input into

the learning machine to construct the feature subspace, and the misclassification rate was calculated by 1NN. This process was repeated 20 times to obtain the average of 20 processes as a result. The error rate versus dimensionality is shown in Figure 3, and the best result obtained in the optimal subspace and the corresponding dimensionality for each method are listed in Table 3. As can be seen, the proposed method performed the best for all the cases. The misclassification of LGSR is lower than that of other methods, and it achieves 90% classification correctness in all dimensions.

Figure 2. The test rig



(a) (b)

Figure 3. The error rate of different algorithms in bearing fault diagnosis: (a) use 30% of the training samples; (b) use 50% of the training samples

Table 3. Performance comparisons on the bearing data

Method	Error rate	
	30% training	50% training
PCA	0.1221	0.0957
LPP	0.1102	0.0855
NPE	0.1008	0.0720
SR	0.1010	0.0695
LGSR	**0.0837**	**0.0599**

4.2. Result Analysis of Case II

Figure 4 shows the vibration signal waveforms of three signals for different defect types of 500, 800, and 1100 RPM. Compared with healthy bearings, it shows a significant difference in the overall vibration amplitude. However, we still need to process the signal to extract the most useful information.

We calculated the original features after the time domain and frequency domain of the above feature extraction method. Figure 5 shows the original features from Table 1. It can be seen that there are some differences in the various defect types of the bearings, but there are still some indistinguishable differences.

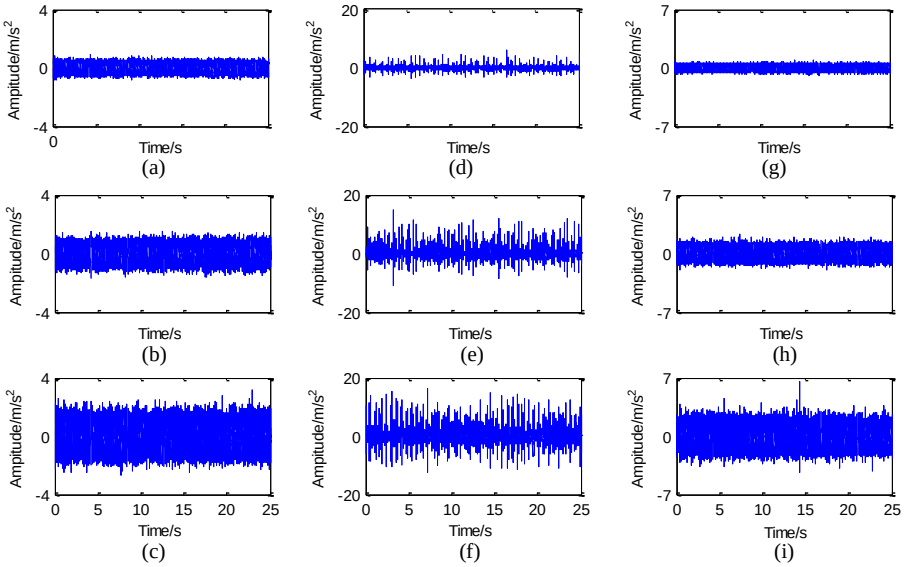
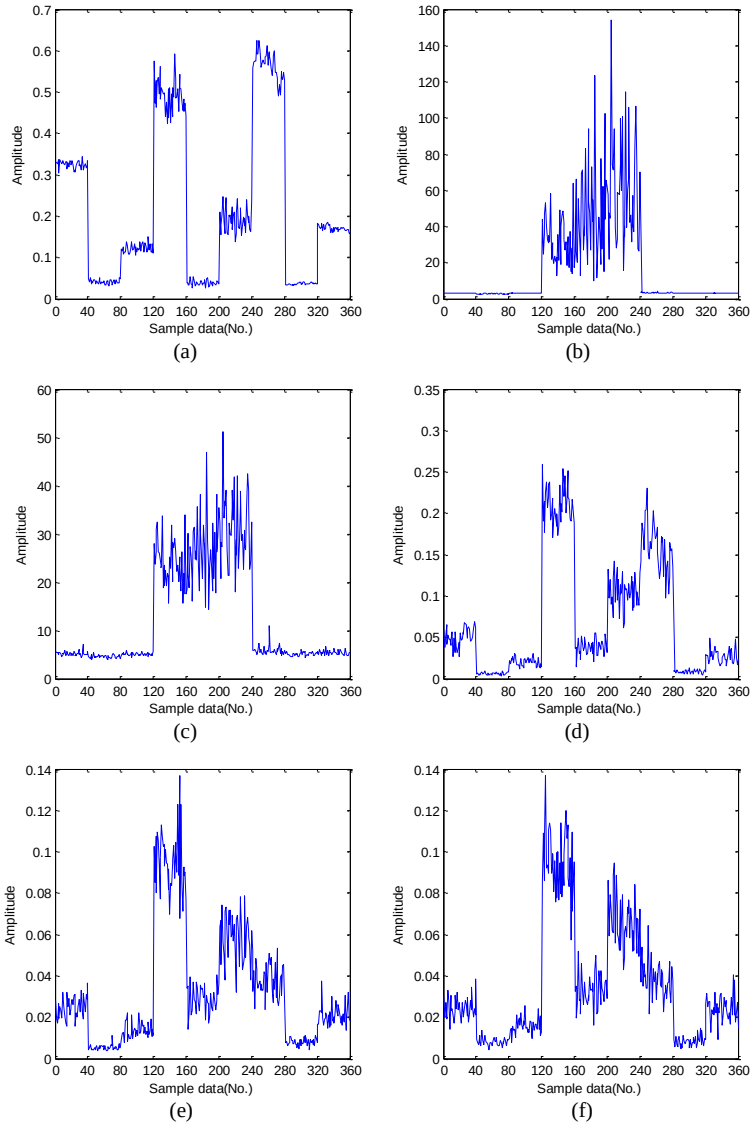
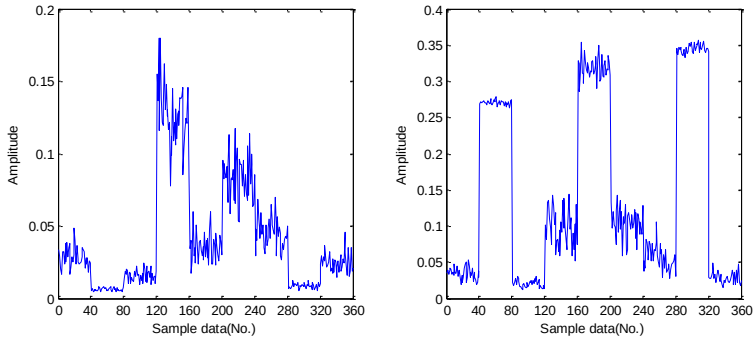


Figure 4. Waveforms of vibration signals from different defect types: (a)-(c) healthy bearings of 500, 800, and 1100 RPM; (d)-(f) inner ring faults of 500, 800, and 1100 RPM; (g)-(i) outer ring faults of 500, 800, and 1100 RPM, respectively

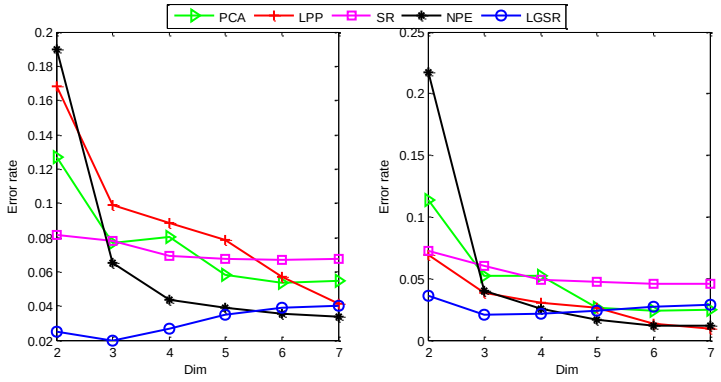




(g) (h)

Figure 5. The time domain and frequency domain characteristics: (a) RMS; (b) kurtosis value; (c) impulse factor; (d) amplitude at rotation frequency; (e) amplitude at BPFI; (f) amplitude at BPFO; (g) amplitude at FTF; (h) amplitude at BSF (Note: Sample data No. 1-40, 41-80, 81-120, 121-160,161-200, 201-240, 241-280, 281－320, and 321-360 represent Normal, Inner fault, and Outer race fault at 500, 800, and 1100 RPM, respectively)

In order to verify the generalization performance of the proposed model, two random subsets with 25% and 50% of the data were used for training, and the rest were used for testing. This process was repeated 20 times, and the average of the 20 processes was the result. The relationship between the error rate and dimension is shown in Figure 6.

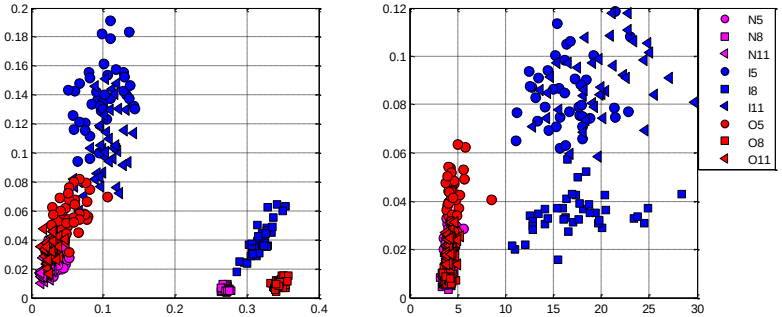


(a) (b)

Figure 6. The error rates for different algorithms in bearing fault diagnosis: (a) use 25% of training samples; (b) use 50% of training samples

As shown in Figure 6, in addition to LGSR, the misclassification rate decreases as the dimension increases, and the training samples increase. When the dimension is greater than 2 (using 50% training samples), all error rates are less than 10%. The performance of LGSR is satisfactory and acceptable. LGSR's misclassification achieved the best classification results in 2-D and 3-D space with error rates from 2.000% to 4.019% (using 25% training samples) and 2.028% to 3.639% (using 50% training samples).

Therefore, we performed two experiments, each of which randomly selected two features from the original feature space, as shown in Figures 7 (a-b). In order to verify the performance of the LGSR method, two features were extracted based on LGSR, as shown in Figure 8 (a). For comparison of different methods, the projected results using PCA, LPP, NPE, and SR are also described in Figures 8 (b-e), respectively.



(a) (b)

Figure 7. Two features randomly selected from the original feature space: (a) the first selection; (b) the second selection

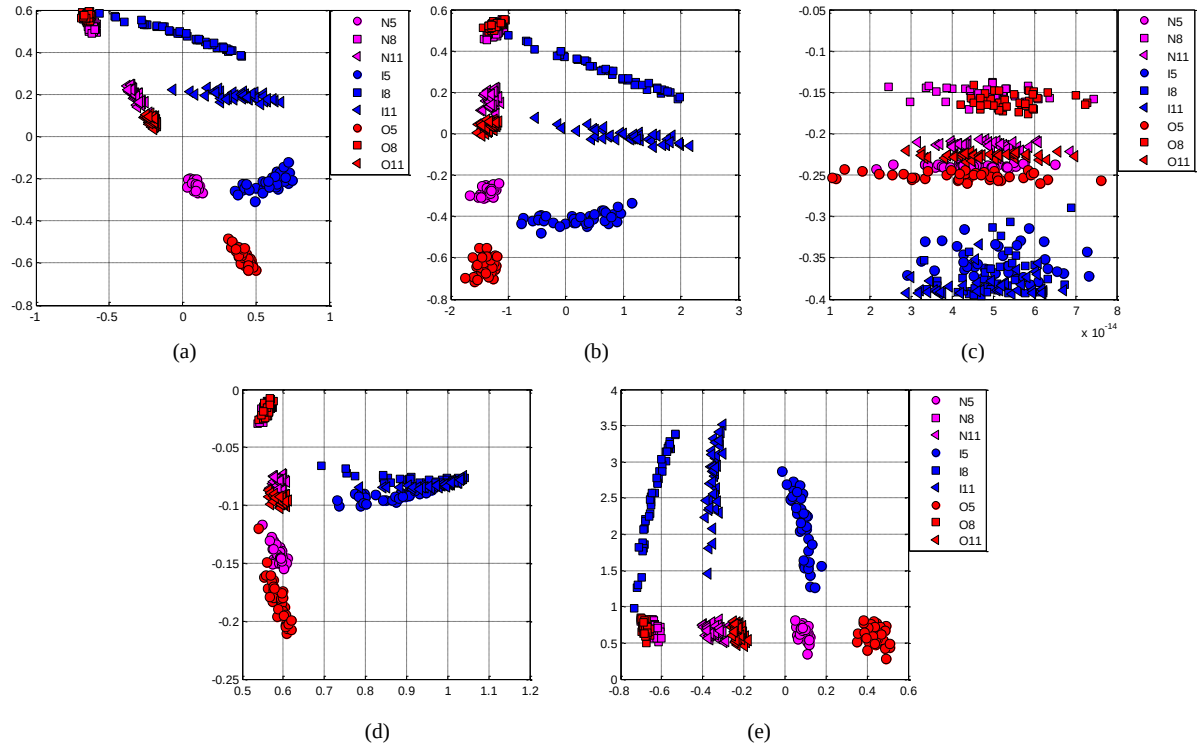


Figure 8. The data projection results of the first two eigenvectors: (a) LGSR; (b) PCA; (c) LPP; (d) NPE; (e) SR (N5: Normal bearing at 500 RPM, N8: Normal bearing at 800 RPM, N11: Normal bearing at 1100 RPM, I5: Inner race fault at 500 RPM, I8: Inner race fault at 800 RPM, I11: Inner race fault at 1100 RPM, O5: Outer race fault at 500 RPM, O8: Outer race fault at 800 RPM, O11: Outer race fault at 1100 RPM)

By comparing the results above, we can observe that LGSR and SR achieve a better clustering effect than the LPP, NPE, and LPPs method. In other words, LGSR and SR have better projection performance compared with the LPP, NPE, and LPP methods. This is because SR can discover locally structured information about data manifolds. In addition, LGSR shows better results than SR, since SR only considers the local structured information and may miss some useful knowledge. However, LGSR aims to discover both local and global information, and it makes the same label of samples become closer while simultaneously making the different labels of samples more distant. Therefore, using the LGSR method to extract features can more effectively improve the performance of classification, further demonstrating that LGSR can extract the most useful information from the original feature space.

5. Conclusions

In order to retain the most useful features, a novel feature extraction algorithm named LGSR is proposed for bearing accelerometer sensor signals. It combines the spectral graph analysis (include local and global spectral graphs) and regression to extract the useful information from original features for regularized subspace learning problems. 1NN is adopted to evaluate the performance of the proposed feature extraction methods, and the bearing fault experiment demonstrates that LGSR can obtain higher accuracy than other similar approaches. The results of cluster experiments show that LGSR can improve the performance of the classification.

Acknowledgements

This work was supported by the Natural Science Foundation of China (No. 51605406, 51605405), Education and Scientific Research Project for Young and Middle-aged Teachers of Fujian (No. JAT170413), Natural Science Foundation of Fujian Province (No. 2018J01531), Research Start-up Funds of DGUT (No. GC300501-26), and Postdoctoral Science Foundation of China (No. 2019M652881).

References

1. S. Dong, Z. Wu, Y. Pan, H. Su, and Y. Liu, "Hidden-Markov-Model-based Asynchronous Filter Design of Nonlinear Markov Jump Systems in Continuous-Time Domain," *IEEE Transactions on Cybernetics*, Vol. 49, No. 6, pp. 2294-2304, 2018

2. L. Wang, Y. Shao, and Z. Cao, "Optimal Demodulation Subband Selection for Sun Gear Crack Fault Diagnosis in Planetary Gearbox," *Measurement*, Vol. 125, pp. 554-563, 2018
3. D. Wang, Y. Zhao, C. Yi, K. L. Tsui, and J. H. Lin, "Sparsity Guided Empirical Wavelet Transform for Fault Diagnosis of Rolling Element Bearings," *Mechanical Systems and Signal Processing*, Vol. 101, pp. 292-308, 2018
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, pp. 673-692, 1997
5. Y. Aït-Sahalia and D. Xiu, "Principal Component Analysis of High-Frequency Data," *Journal of the American Statistical Association*, Vol. 114, No. 525, pp. 287-303, 2019
6. R. O. Duda and P. E. Hart, "Pattern Classification and Scene Analysis," John Wiley and Sons, New York, 1973
7. R. Q. Yan and R. X. Gao, "Approximate Entropy as a Diagnostic Tool for Machine Health Monitoring," *Mechanical Systems and Signal Processing*, Vol. 21, pp. 824-839, 2007
8. J. B. Yu, "Local and Nonlocal Preserving Projection for Bearing Defect Classification and Performance Assessment," *IEEE Transactions on Industrial Electronics*, Vol. 59, pp. 2363-2376, 2012
9. Y. Wang, G. Xu, L. Liang, and K. Jiang, "Detection of Weak Transient Signals based on Wavelet Packet Transform and Manifold Learning for Rolling Element Bearing Fault Diagnosis," *Mechanical Systems and Signal Processing*, Vol. 54, pp. 259-276, 2015
10. X. Ding, Q. Li, L. Lin, Q. He, and Y. Shao, "Fast Time-Frequency Manifold Learning and Its Reconstruction for Transient Feature Extraction in Rotating Machinery Fault Diagnosis," *Measurement*, Vol. 141, pp. 380-395, 2019
11. X. Ding, Q. He, Y. Shao, and W. Huang, "Transient Feature Extraction based on Time-Frequency Manifold Image Synthesis for Machinery Fault Diagnosis," *IEEE Transactions on Instrumentation and Measurement*, Vol. 68, No. 11, pp. 4242-4252, 2019
12. D. Cai, "Spectral Regression: A Regression Framework for Efficient Regularized Subspace Learning," Ph.D. Thesis, University of Illinois at Urbana-Champaign, Urbana, IL, USA, 2009
13. Z. G. Xia, S. X. Xia, L. Wan, and S, Y. Cai, "Spectral Regression based Fault Feature Extraction for Bearing Accelerometer Sensor Signals," *Sensors*, Vol. 12, pp. 13694-13719, 2012
14. W. H. Li, T. L. Shi, G. L. Liao, and S. Z. Yang, "Feature Extraction and Classification of Gear Faults using Principal Component Analysis," *Journal of Quality in Maintenance Engineering*, Vol. 9, pp. 132-143, 2003
15. P. Tse and D. Atherton, "Prediction of Machine Deterioration using Vibration based Fault Trends and Recurrent Neural Networks," *Journal of Vibration and Acoustics-Transactions of the ASME*, Vol. 121, pp. 355-362, 1999
16. K. Ding, W. Li, and X. Zhu, "Practical Technologies for Gear and Gearbox Fault Diagnosis," China Machine Press, Beijing, August, 2005
17. K. A. Loparo, et al., "Bearingdata Center, Case Western Reserve University," (http://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website, 2009)