

Article

A Semi-Supervised Approach to Bearing Fault Diagnosis under Variable Conditions towards Imbalanced Unlabeled Data

Xinan Chen ^{1,2,3}, Zhipeng Wang ^{1,2,3,*} , Zhe Zhang ^{1,2,3}, Limin Jia ^{1,2,3,*} and Yong Qin ^{1,2,3}

¹ State Key Lab of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China; 15114217@bjtu.edu.cn (X.C.); zhangzhe@bjtu.edu.cn (Z.Z.); yqin@bjtu.edu.cn (Y.Q.)

² National Engineering Laboratory for System Safety and Operation Assurance of Urban Rail Transit, Guangzhou 510000, China

³ Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Beijing 100044, China

* Correspondence: zpwang@bjtu.edu.cn (Z.W.); lmjia@bjtu.edu.cn (L.J.); Tel.: +86-010-5168-4281 (Z.W.); +86-010-5168-4639 (L.J.)

Received: 27 May 2018; Accepted: 27 June 2018; Published: 29 June 2018



Abstract: Fault diagnosis of rolling element bearings is an effective technology to ensure the steadiness of rotating machineries. Most of the existing fault diagnosis algorithms are supervised methods and generally require sufficient labeled data for training. However, the acquisition of labeled samples is often laborious and costly in practice, whereas there are abundant unlabeled samples which also imply health information of bearings. Thus, it is worthwhile to develop semi-supervised methods of fault diagnosis to make effective use of the plentiful unlabeled samples. Nevertheless, considering the normal data are much more than the faulty ones, the problem of imbalanced data exists among unlabeled samples for fault diagnosis. Besides, in practice, bearings often work under uncertain and variable operation conditions, which would also have negative influence on fault diagnosis. To solve these issues, a novel hybrid method for bearing fault diagnosis is proposed in this paper: (1) Inspired by visibility graph, a novel fault feature extraction method named visibility graph feature (VGF) is proposed. The obtained features by VGF are natively insensitive to variable conditions, which has been validated by a simulation experiment in this paper; (2) On basis of VGF, to deal with imbalanced unlabeled data, graph-based rebalance semi-supervised learning (GRSSL) for fault diagnosis is proposed. In GRSSL, a graph based on a weighted sparse adjacency matrix is constructed by the k-nearest neighbors and Gaussian Kernel weighting algorithm by means of the samples. Then, a bivariate cost function over classification and normalized label variable is built up to rebalance the importance of labels. Finally, the proposed VGF-GRSSL method was verified by data collected from Case Western Reserve University Bearing Data Center. The experiment results show that the proposed method of bearing fault diagnosis performs effectively to deal with the imbalanced unlabeled data under variable conditions.

Keywords: rolling element bearing; semi-supervised learning; visibility graph; imbalanced data

1. Introduction

Rolling element bearings are one of the most frequently used components in machines. Reference [1] shows that a large part of faults in machines (above 40% of the total faults) owes to bearings. Therefore, bearing fault diagnosis has been a research hotspot for decades. Reference [2] reviewed the diagnostic techniques over the past decade and summarized six trends in fault diagnosis for the electrical machines. So far, many studies have proposed plenty of feature extraction

methods, such as Fourier transform, wavelet transform, empirical mode decomposition, fuzzy entropy et al. [3–6]. Besides, machine learning methods, including support vector machine, decision tree, neural network [7–9], are utilized to classify various bearing fault categories and achieve high accuracy rates of fault diagnosis. In addition, many studies have used novel observation of the bearing to detect incipient phase bearing faults, such as those based on current signals and stray flux measurement in different positions [10,11].

However, although the aforementioned algorithms perform well, reference [12] indicates that the classifications based on supervised learnings tend to perform poorly due to inadequate labeled training data. It is because the classifiers are supposed to remember the training samples instead of learning rules from them and easily lead to overfitting. However, this issue is rarely discussed in the field of mechanical fault diagnosis. In practice, it is laborious or expensive to collect faulty samples with labels, whereas the unlabeled samples are abundant [13,14]. Therefore, it is valuable to develop semi-supervised methods for fault diagnosis to improve the accuracy as much as possible by means of a few labeled data and mass unlabeled data.

Meanwhile, the monitoring data of bearings are usually collected during runtime. Since most of the time, bearings work under normal conditions, normal samples acquired from bearings are much more than the faulty ones. Therefore, the distributions of bearing data are seriously imbalanced in practice [15]. Therefore, the semi-supervised approach to bearing fault diagnosis suffers from the problem of imbalanced data. Besides, in the real world, bearings often work under variable and fluctuant conditions. The labeled and unlabeled data are gathered from bearings with speed variations in practice. It is essential to study feature extraction under variable conditions. Therefore, to deal with these aforementioned issues, this paper mainly discusses how to develop a highly-accurate diagnosis for bearings under variable conditions with imbalanced unlabeled data.

To this end, a novel hybrid method for bearing fault diagnosis is presented in this paper. Inspired by visibility graph, a novel fault feature extraction method named visibility graph feature (VGF) is proposed to extract fault characteristics under variable conditions. The VGF method converts vibration signals into graphs and extracts features based on structures of these graphs. Since the visibility remains invariant under horizontal and vertical transformation of time series, the obtained VGFs are natively insensitive to variable conditions, which is verified in the following simulation experiments. Based on the VGFs, graph-based rebalance semi-supervised learning (GRSSL) is employed for fault classification. As a new semi-supervised method, GRSSL [16] is proposed by in. In GRSSL, a graph is established from data based on a kernel function, and then the k-nearest neighbors (KNN) and Gaussian Kernel weighting algorithm are employed to sparsify and reweight the connections. During the optimization process, a bivariate function over classification function and the unknown labels is built up, and then, a normalized label variable is used to rebalance the importance of labels. In this paper, experiments are designed to test the performance of the proposed GRSSL and VGF. The results have shown that the GRSSL outperforms the popular methods, including Gaussian Fields and Harmonic Functions (GFHF) [17] and Local and Global Consistency (LGC) [18].

The rest of the paper is organized as follows. Section 2 reviews the related literature. The novel VGF and GRSSL are presented in Section 3. Section 4 describes the simulation experiment for VGF. Experiments are analyzed in detail in Section 5, and Section 6 presents the key conclusions.

2. Related Work

Feature extraction is a key step for bearing fault diagnosis. There have been many studies on feature extraction under variable conditions. Borghesani et al. [19] proposed a new procedure for using envelop analysis to remove the effects of variable conditions, but required prior knowledge about defect frequencies of bearings. Feng et al. [20] exploited a concentration of frequency and time method to deal with the variable speed conditions, but required prior information about the undergoing conditions. Tian et al. [21] employed local mean decomposition method and extreme learning machine to detect the bearing fault under variable operation conditions, but required complete condition data.

Wang et al. [22] conjugated variation mode decomposition and singular value decomposition to extract features which can fit the variable conditions adaptively. However, it is even impossible for such prior knowledge and complete data for all conditions. Therefore, it is essential to extract fault features which are natively insensitive to variable conditions. Visibility graph(VG), proposed by Lacasa et al. [23] intends to convert time series into complex networks, and the obtained structure parameters not only imply the characteristics of the raw data, but also keep invariant under several transformations of the data. Therefore, inspired by VG, a novel feature extraction method under variable conditions is proposed in this paper.

Fault classification is the following step after feature extraction. There are a large number of references on vibration-based bearing fault classification [7–9]. However, most of the existing methods purely depend on labeled data for training. The masses of unlabeled samples acquired in practice are abandoned. To make effective use of those valuable unlabeled samples, the semi-supervised learning (SSL) [24], which is capable of exploiting the unlabeled data combined with a small amount of labeled data to train a well-performed classifier, has been a highlight issue in the field of bearing fault diagnosis.

Concerning this issue, Li [25] proposed a semi-supervised weighted kernel clustering algorithm based on gravitational search for bearing fault diagnosis and processed the unlabeled samples by calculating the weighted kernel distances among them and fault cluster centers. Qin et al. [26] employed a tritraining method for bearing fault diagnosis. Zhao et al. [27] proposed a new graph-based semi-supervised classification for bearing fault diagnosis with sparse coding method. However, these existing semi-supervised methods of bearing fault diagnosis have not considered the imbalance of the unlabeled data. Therefore, it is crucial to develop novel semi-supervised methods for bearing fault diagnosis to deal with the imbalanced unlabeled data.

Among the semi-supervised methods, Graph-based semi-supervised learning (GSSL) [17,18] considers samples as vertices in a graph and shapes the pairwise edges, which are determined by the similarity between the corresponding samples. Then, the small portion of labels propagate to predict labels for unlabeled samples by semi-supervised learning method. Subramanya et al. [28] have proved that GSSL outperforms non-graph-based SSL approaches. Based on the GSSL, Jebara et al. [16] extends a bivariate optimization by which both classification function and a normalized label variable are considered to rebalance the importance of labels. Therefore, the graph-based rebalance semi-supervised learning (GRSSL) is employed for fault classification to deal with the imbalanced unlabeled data.

The intended contributions of this study can be summarized as follows:

1. A novel fault feature extraction method named visibility graph feature (VGF) is proposed to extract fault characteristics under variable conditions. The vibration signals are mapped into graphs whose structures are natively insensitive to variations of signals corresponding to the speed and load conditions.
2. A GRSSL algorithm by blending the normalized label variable is developed for bearing fault diagnosis with imbalanced and unlabeled data. To cope with the imbalanced problem, a bivariate optimization function is employed, and the importance of labels is rebalanced by the normalized label variable term.

3. Methodology

In this section, the proposed fault diagnosis method based on VGF-GRSSL is described in detail, as shown in Figure 1. There are four layers in the method: data segmentation layer, feature extraction layer, fault classification layer, and output layer.

1. Data segmentation layer: The signals are segmented according to the sample rate and rough shaft speed to ensure each obtained sample covers several circles of signals.

2. Feature extraction layer: The samples are mapped into visibility graphs. Then, features based on the structures of graphs including degree distribution of visibility graph (DDVG) and graph index complexity (GIC) are extracted.
3. Fault classification layer: This layer includes graph construction and label propagation.

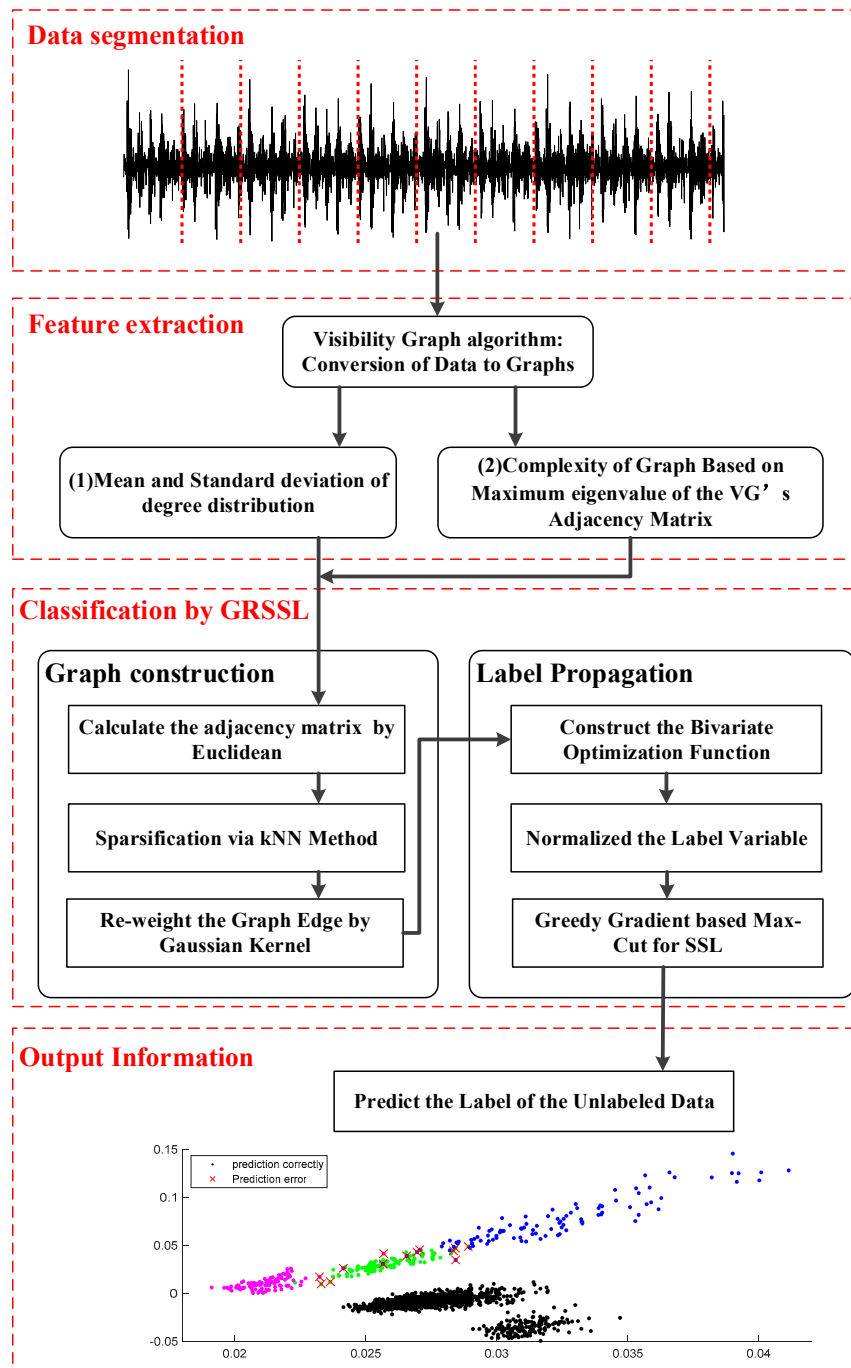


Figure 1. The overall scheme of proposed fault diagnosis method.

In the graph construction, an adjacency matrix is calculated based on a kernel function firstly, and then sparsified and reweighted by means of the k-nearest neighbors (KNN) and Gaussian Kernel weighting algorithm, respectively. Then, a graph with a weighted sparse undirected adjacency matrix is calculated from the samples.

During the label propagation process, a bivariate function over classification function and the unknown labels is built up, and then, a normalized label variable is used to rebalance the importance of labels. Finally, the greedy gradient-based Max-Cut algorithm is adopted to solve the optimization model.

4. Output layer: The unlabeled data is marked with the corresponding label according to the classification with few labeled data. The GRSSL classification based on VGF is expected to extend the bearing fault diagnosis under variable condition with imbalanced unlabeled data.

3.1. Visibility Graph Feature Extraction

3.1.1. Construction of Visibility Graph

The theory of visibility graph has been lucubrated for years and applied widely on many fields, such as engineering and urban planning [29]. Lacasa et al. [23] introduced the theory to the analysis of time series. Here, an undirected complex network is established by individual observations in a time series whose connectivity is defined through the visibility condition in physical space. It has been proved that the obtained graph inherits the intrinsic properties of the raw time series. For instance, the values of impulses of time series are related to the hubs of the graph, and the periods of time series are related to the degree distributions of the graphs. Motivated by this creative idea, this paper proposes a novel feature extraction method based on visibility graph.

An example is given to describe how to convert a time series into a visibility graph, as shown in Figure 2. There are 10 points of a random series. Then, the visibility exists between any two points if there is a straight line that does not intersect any intermediate data height. The complexity of the time series is supposed to be revealed in the structure of the obtained visibility graph.

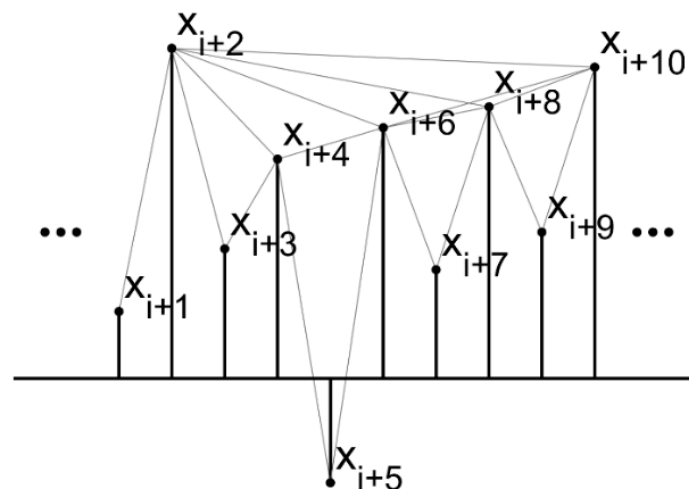


Figure 2. Illustration of converting a time series into a visibility graph.

The visibility criteria can be formulated as follows: two arbitrary points (t_a, x_a) and (t_b, x_b) of the series data will be connected as nodes of the graph, if the points (t_c, x_c) between them fulfill:

$$x_c < x_b + (x_a - x_b) \frac{t_b - t_c}{t_b - t_a} \quad (1)$$

It is obviously that the associated graph constructed from a time series is always:

1. Connected: each node can see its nearest two neighbors at least.
2. It is an undirected link between every two node.

3. Invariant under affine transformations: the visibility keeps invariant under the resizing of both horizontal and vertical axes (as shown in Figure 3). Therefore, the visibility-graph-based features are natively insensitive to variable conditions.

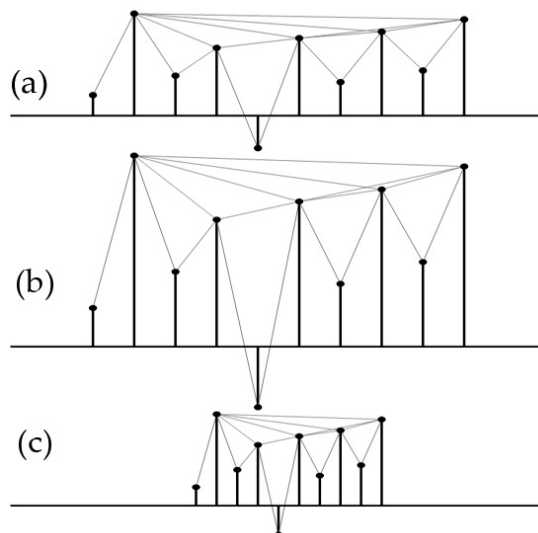


Figure 3. The visibility remains invariant under horizontal and vertical transformation of the time series. (a) Visibility links of the original time series; (b) Resizing vertically; (c) Resizing horizontally.

3.1.2. Features Based on Visibility Graph

After the construction of the graph, an adjacent matrix $W_V \in B^{n \times n}$ can be defined to record the links of the graph. It is a symmetric matrix, and the values of elements are either 1 or 0, which indicate that there is a connection or not between the corresponding nodes. For example, if one of the elements $w_{ij} = 1$, it indicates that the node i and node j are connected with each other. Then, degree distribution of visibility graph (DDVG) and graph index complexity (GIC) are involved for feature extraction under variable conditions.

Degree Distribution of Visibility Graph (DDVG)

Define the degree distribution $D_V = [d_1, \dots, d_n]$ with $d_i = \sum_{j=1}^n w_{ij}$, which indicate the connection situation of the graph. Then, the mean and standard deviation of the degree distribution ($Mean(D_v)$ and $Std(D_v)$) are calculated as fault features.

Figure 4a shows a sample of faulty data with 1024 points. Figure 4b shows the degree distribution of its VG with 1024 nodes. There are several nodes with large values in the degree distribution. Such nodes are hubs of the graph and imply the corresponding points in the time series have large or small values.

Graph Index Complexity (GIC)

As a measure of complexity of a graph, GIC [30] is also involved for feature extraction in this study. GIC is defined as follows:

$$C = 4c(1 - c) \quad (2)$$

where

$$c = \frac{\lambda_{\max} - 2\cos(\pi/(n+1))}{n-1-2\cos(\pi/(n+1))} \quad (3)$$

λ_{\max} represents the largest eigenvalue of the adjacency matrix of a graph with n nodes.

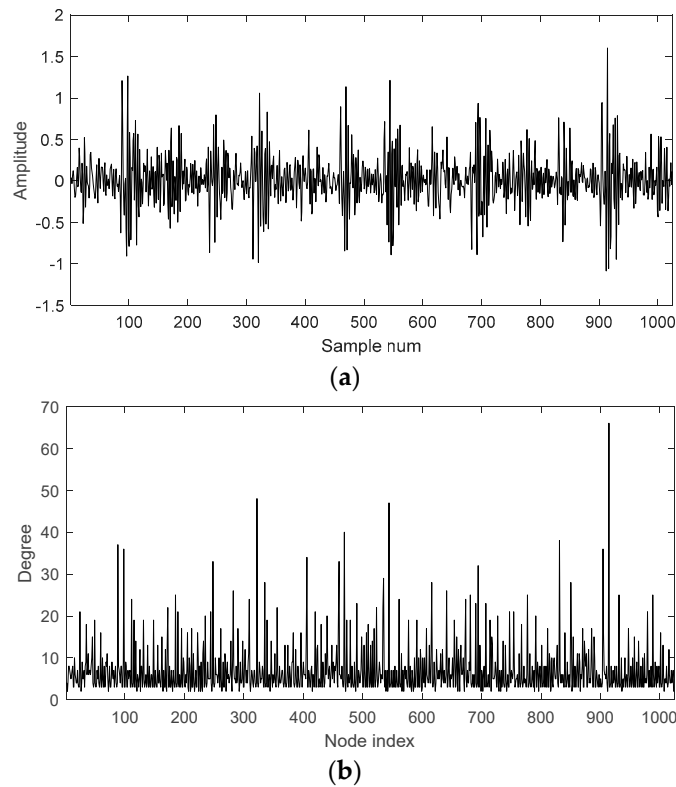


Figure 4. (a) A sample of faulty data and (b) degree distribution of its VG.

3.2. Fault Classification Based on GRSSL

Assume that there are labeled samples $\{(x_1, y_1), \dots, (x_l, y_l)\}$ and unlabeled samples $\{x_{l+1}, \dots, x_{l+u}\}$ derived from an identical independent distribution. Define $X_l = \{x_1, \dots, x_l\}$ and $X_u = \{x_{l+1}, \dots, x_{l+u}\}$ as the set of labeled inputs and unlabeled inputs, respectively. l and u represent the number of labeled and unlabeled samples. $Y = \{y_{ij}\} \in \mathbb{B}^{n \times c}$ is the label information matrix, where $y_{ij} = 1$ if the sample x_i is related to label j , $j \in \{1, 2, \dots, c\}$. Each sample can be marked with a label: $\sum_{j=1}^c y_{ij} = 1$. The semi-supervised learning aims to identify the missing labels $\{y_{l+1}, \dots, y_{l+u}\}$ of unlabeled samples, where typically $l \ll n$ ($l + u = n$); there are two parts of the graph-based rebalance semi-supervised learning (GRSSL): graph construction and label propagation.

3.2.1. Graph Construction

In this paper, assume the $G = \{X, E, W\}$ represents the undirected graph from the data $X = X_l \cup X_u$. There are usually two steps in the estimation of G from X .

1. The sample similarities are calculated by using a kernel function and utilized to construct a full adjacency matrix K : $K \in \mathbb{R}^{n \times n}$, $K_{ij} = k(x_i, x_j)$
2. The matrix K is c and reweighted to obtain the final matrix W .

The sparsification deletes edges of the matrix K by multiplying a binary matrix $B \in \mathbb{B}^{n \times n}$ and a distance matrix $H \in \mathbb{R}^{n \times n}$. The binary and distance matrix are defined as follows:

$$B_{ij} = \begin{cases} 1 & x_i \text{ connect } x_j \\ 0 & x_i \text{ disconnect } x_j \end{cases} \quad (4)$$

$$H_{ij} = \sqrt{K_{ii} + K_{jj} - 2K_{ij}} \quad (5)$$

In this study, the k -nearest neighbor algorithm is employed to estimate the binary matrix B . The optimization procedure can be defined as follows:

$$\begin{aligned} \min_{B \in B} \sum_{ij} \hat{B}_{ij} H_{ij} \\ \text{s.t. } \sum_j \hat{B}_{ij} = k, \hat{B}_{ii} = 0, \forall i, j \in 1, \dots, \mathbf{n}. \end{aligned} \tag{6}$$

The final binary matrix is computed by:

$$B_{ij} = \mathbf{max}(\hat{B}_{ij}, \hat{B}_{ji}) \tag{7}$$

After the construction of the sparsified graph, the Gaussian Kernel Weighting algorithm is employed to calculate the weight matrix W . The weight between x_i and x_j is computed as follows:

$$w_{ij} = B_{ij} \mathbf{exp} \left(-\frac{d^2(x_i, x_j)}{2\sigma^2} \right) \tag{8}$$

where, $d(x_i, x_j)$ represent the Euclidean distance of x_i and x_j .

Finally, a graph G with a weighted sparse adjacency matrix W is generated from the data.

3.2.2. Label Propagation

Given the constructed graph $G = \{X, E, W\}$, the purpose of label propagation is to diffuse the known labels to all unlabeled nodes X_u in the graph and infer Y_u . In this study, this issue is cast as a bivariate optimization over the classification function F and the unknown labels Y :

$$\begin{aligned} (F^*, Y^*) &= \mathbf{arg\ min}_{F \in R^{n \times c}, Y \in B^{n \times c}} Q(F, Y) \\ &= \mathbf{arg\ min}_{F \in R^{n \times c}, Y \in B^{n \times c}} (Q_{smooth}(F) + Q_{fit}(F, Y)) \\ \text{s.t. } &y_{ij} \in \{0, 1\}, \\ &\sum_{j=1}^c y_{ij} = 1, j = 1, \dots, c. \\ &y_{ij} = 1, \text{ for } label(x_i) = j, j = 1, \dots, c. \end{aligned} \tag{9}$$

where Y is a binary matrix. The constraint $\sum_{j=1}^c y_{ij} = 1$ indicates that it is a single label prediction issue, and (F^*, Y^*) represents the optimal solution. The cost function can be specified as:

$$\begin{aligned} Q(F, Y) &= \|F\|_G^2 + \frac{\mu}{2} \|F - Y\|^2 \\ &= \frac{1}{2} \text{tr} \left(F^T L F + \mu (F - Y)^T (F - Y) \right) \end{aligned} \tag{10}$$

where the first item ranks the function smoothness [31] over graph G , and the second item represents the loss to fit label matrix. The normalized graph Laplacian are as follows:

$$L = D^{-1/2} (D - W) D^{-1/2} = I - D^{-1/2} W D^{-1/2} \tag{11}$$

where the vertex degree matrix $D = \text{diag}([d_1, \dots, d_n])$ with $d_i = \sum_{j=1}^n w_{ij}$.

The cost function can be rewritten as follows:

$$Q(F, Y) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} \left\| \frac{F_{i \cdot}}{\sqrt{d_i}} - \frac{F_{j \cdot}}{\sqrt{d_j}} \right\|^2 + \frac{\mu}{2} \sum_{i=1}^n \|F_{i \cdot} - Y_{i \cdot}\|^2 \tag{12}$$

where $F_{i \cdot}$ denote the i 'th row vectors of F .

The bivariate issue is converted to a univariate issue in regards to the label variable Y .

(1) Optimization of F :

F is continuous, and the issue is convex while Y is fixed. The minimum is easy to be calculated by setting the partial derivative to zero:

$$\begin{aligned} \frac{\partial Q}{\partial F^*} = 0 &\Rightarrow LF^* + \mu(F^* - Y) = 0 \\ \Rightarrow F^* &= (L/\mu + \mathbf{I})^{-1}Y = PY \end{aligned} \quad (13)$$

where, $P = (L/\mu + \mathbf{I})^{-1}$ is defined as the propagation matrix.

(2) Optimization of Y :

The optimal F^* is employed in Equation (10) instead of F .

$$\begin{aligned} Q(Y) &= \frac{1}{2}tr(Y^T P^T L Y P + \mu(PY - Y)^T (PY - Y)) \\ &= \frac{1}{2}tr(Y^T (P^T L P + \mu(P^T - I)(P - I))Y) = \frac{1}{2}tr(Y^T A Y) \end{aligned} \quad (14)$$

where, $A = P^T L P + \mu(P^T - I)(P - I) = P^T L P + \mu(P - I)^2$.

To cope with the imbalanced data, a normalized label variable $\tilde{Y} = \Lambda Y$ is utilized to replace the original Y . The diagonal matrix $\Lambda = \text{diag}([\lambda_1, \dots, \lambda_n])$ is imported to rebalance the importance of labels based on node degrees. In the formula (15), the degree of the vertex represents its importance in the graph. The majority class is supposed to correspond to a larger $\sum_k y_{kj} d_k$ than that of the minority class, so the sample from majority has a small λ_i that can reduce its importance in the graph. The sum of λ_i of each class is equal to 1. That means that they shared the same importance in the graph. Therefore, the diagonal matrix is expected to eliminate the influence of the imbalanced data.

$$\lambda_i = \begin{cases} p_j \cdot \frac{d_i}{\sum_k y_{kj} d_k} & y_{ij} = 1 \\ 0 & otherwise \end{cases} \quad (15)$$

where d is the degree of the vertex, p is the prior distribution of the class and constrained to $\sum_{j=1}^c p_j = 1$.

In this paper, $p_j = 1/c$ by default.

The final optimization issue is rewritten as follows:

$$\begin{aligned} Y^* &= \mathbf{arg\,min} \frac{1}{2}tr(Y^T \Lambda A \Lambda Y) \\ \text{s.t. } & y_{ij} \in \{0, 1\}, \\ & \sum_{j=1}^c y_{ij} = 1, j = 1, \dots, c. \\ & y_{ij} = 1, \text{ for } \text{label}(x_i) = j, j = 1, \dots, c. \end{aligned} \quad (16)$$

The optimization is a NP hard problem which requires solving a linearly constrained binary integer programming (BIP) problem [32]. A forthright way to work out the minimization problem is to update the label variable Y with the gradient descent method. Wang and Jebara [16] has proved that the minimization problem equals to a Max K-Cut problem ($K = c$) over the graph $G_A = \{X, A\}$.

Therefore, this study utilizes a greedy gradient Max-Cut algorithm to find local optima by selecting unlabeled vertices randomly and placing each of them into the appropriate class subset with minimum connectivity to maximize cross-set edge weights iteratively. The connectivity between unlabeled vertex x_i and labeled subset S are defined as follows:

$$c_{ij} = p_j \cdot \sum_{m=1}^n \lambda_m a_{im} y_{mj} = p_j \cdot \Lambda A_i \cdot Y_j \quad (17)$$

$$S_j = \{x_i | y_{ij} = 1\}, i = 1, 2, \dots, n; j = 1, 2, \dots, c \quad (18)$$

The greedy gradient Max-Cut Algorithm 1.

Algorithm 1. Greedy Gradient Max-Cut

```

1  Input: the graph  $G_A = \{X, A\}$  and labeled vertex  $X_l$  and label  $Y$ ;
2  Initialization:
3  Construct the initial cut  $\{S_j\}$  by the initial labeled vertex  $X_l$ :
    $S_j = \{x_i | y_{ij} = 1\}, j = 1, 2, \dots, c$ 
4  Unlabeled vertex set  $X_u = X \setminus X_l$ ;
5  repeat
6    for all  $j = 0$  to  $|X_u|$  do
7      calculate the connectivity:
    $c_{ij} = \sum_{k=1}^n \lambda_k a_{ik} y_{kj}, x_i \in X_u, j = 1, \dots, c$ 
8    End for
9    Update the cut  $\{S_j\}$  by placing the vertex  $x_i^*$  to the  $S_{j^*}$  subset:
    $(i^*, j^*) = \arg \min_{i, j, x_i \in X_u} c_{ij}$ 
10   Add  $x_i$  to  $X_l : X_l \leftarrow X_l + x_i$ ;
11   Delete  $x_i$  from  $X_u : X_u \leftarrow X_u - x_i$ ;
12 Until  $X_u = \emptyset$ 
13 Output: the final cut and the corresponding labeled subsets  $S_j, j = 1, 2, \dots, c$ 

```

4. Simulation Experiment for VGF

To demonstrate the performance of VGF, a faulty simulated rolling bearing vibration signal with different resonant frequency under variable speeds is generated and analyzed. The simulated signal can be generated as follows [33]:

$$x(t) = \sum_{k=-N}^N A e^{-\alpha(t-k*p*f_r - \sum_{i=-N}^k \tau_i)} \sin\left(2\pi f_m \left(t - k * p * f_r - \sum_{i=-N}^k \tau_i\right)\right) u\left(t - k * p * f_r - \sum_{i=-N}^k \tau_i\right) + n(t) \quad (19)$$

where, $\alpha = 0.03$ is the structural damping characteristic, $f_r = n_s/60$ (n_s means the shaft speed) is the shaft rotation frequency, $f_m = 2000$ Hz is the resonant frequency, $A = f_r/20$ is the amplitude of the impulse. p represents the number of fault impulses in every shaft revolution (here $p = 3.58$). τ_i denotes the randomness of rolling elements slippage, which is subject to a uniformly distribution with a zero mean and a standard deviation of $(0.01 \sim 0.02)f_r$. $n(t)$ is a white Gaussian noise with a signal-to-noise ratio of 0 dB.

To simulate variable conditions, this study assumed that the shaft speed varies from 1500 to 1800 r/min, and the step size defaults to 30 r/min. Therefore, 11 groups of simulated signals were gathered with a sample frequency of 12 kHz, as shown in Figure 5.

To verify the VGF, the signals were segmented into samples with a fixed length ($N = 1024$). Then, GIC was extracted from the samples. For comparisons, Approximate Entropy [34], Sample Entropy [35], and Fuzzy Entropy [36] are involved in this study. This paper set the embedding dimension $m = 2$, similarity criterion $r = 0.15$ of the standard deviation of a signal, length of data $N = 1024$, and fuzzy power $n = 2$.

The averages of features are calculated. There are 11 values corresponding to different speeds for each method, as shown in Table 1. It is obvious that GIC has the smallest standard deviation which means it is natively insensitive to variable conditions.

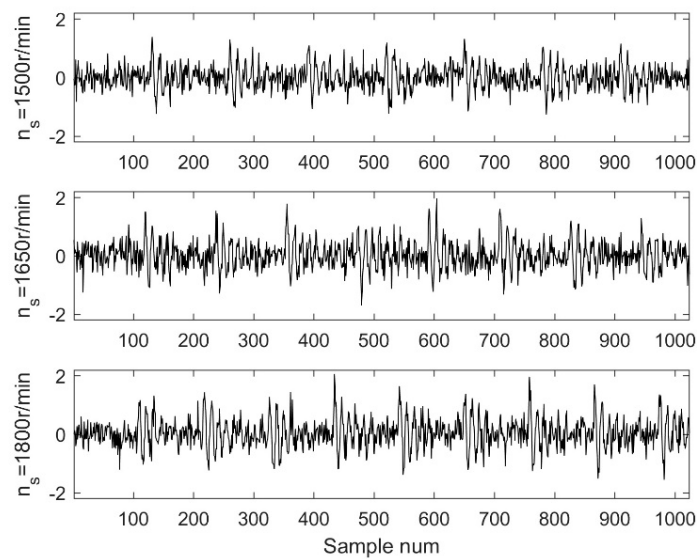


Figure 5. The simulated signal with different n_s .

Table 1. Features of the simulated signal.

Shaft Speed/Feature	GIC	ApEn	SampEn	FuzEn
1500 r/min	0.046	1.485	2.288	0.782
1530 r/min	0.046	1.496	2.313	0.780
1560 r/min	0.046	1.485	2.272	0.776
1590 r/min	0.047	1.496	2.319	0.795
1620 r/min	0.046	1.478	2.268	0.783
1650 r/min	0.047	1.478	2.252	0.786
1680 r/min	0.045	1.472	2.240	0.774
1710 r/min	0.047	1.469	2.209	0.773
1740 r/min	0.047	1.476	2.243	0.786
1770 r/min	0.046	1.458	2.197	0.782
1800 r/min	0.046	1.460	2.184	0.762
Std of features	0.0005	0.0126	0.0444	0.0085

5. Experimental Analysis

5.1. Experimental Setup

This study utilized the data collected by the Case Western Reserve University Bearing Data Center to verify the proposed method. The test rig is shown in Figure 6, the shaft is driven by a 2 hp (1470 W) reliance electric motor, a torque transducer, and an encoder mounted on the shaft (Details about the test rig and experimental data can be found in [37]). The tested bearings are 6205-2RS JEM SKF, deep groove ball bearings. The faults were introduced using electro-discharge machining ranging from 0.007 inches (0.178 mm) in diameter to 0.040 inches (0.355 mm) in diameter separately at the inner race, rolling element, and outer race. The location of faults at the fixed outer race was considered with 3 o'clock, 6 o'clock and 12 o'clock. With the motor loads ranging from 0 to 3 horsepower (0 to 2205 W) (motor speeds of 1797 to 1720 r/min), vibration data was recorded at sample frequencies of 12 kHz and 48 kHz.

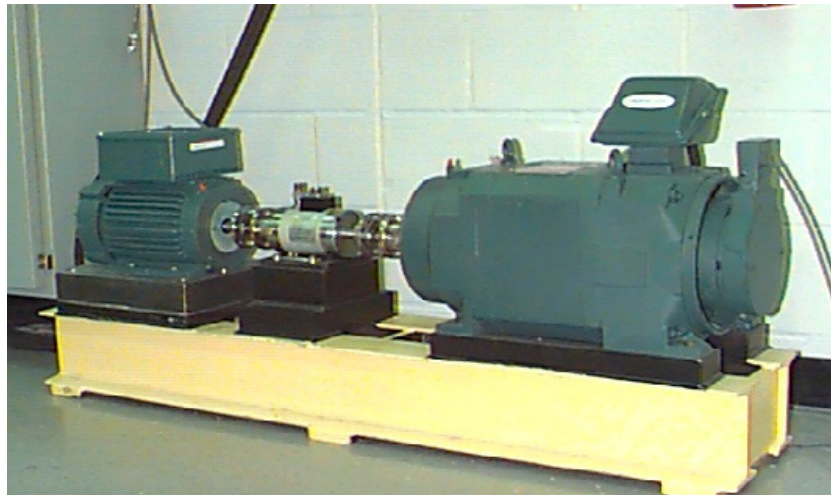


Figure 6. Test-rig of the rolling bearing [37].

To demonstrate the superiority of the proposed VGF-GRSSL method, this study randomly selected a set of imbalanced data with unlabeled samples under variable conditions for comparison and analysis. The dataset under the defect of 0.007 inches (0.178 mm) and sample frequencies of 12 kHz is shown in the Table 2. The vibration data was divided into four types including normal, inner race fault, outer race fault, and rolling element fault. All of them consist of four operating conditions with the load range from 0 to 3 horsepower (0 to 2205 W) and speed range from 1797 to 1720 r/min, respectively. Each sample contains 1024 points. Therefore, there are 3081 samples (1657 for normal, 476 for inner fault, 475 for outer fault, and 473 for element fault). Generally speaking, the sampling time for each data sample is extremely short. For example, when the sample frequency is set to 12 kHz and the length of each data sample $N = 1024$, the sampling time is less than 0.085 s. During such a short time period, the operation condition is considered as constant. Therefore, the four different operation conditions of the Case Western data can be considered under variable conditions.

Table 2. Dataset of 0.007 inches (0.178 mm) defect.

Dataset	Fault Type	Operating Condition	Motor Load (hp)	Sample Num
1	Normal	1797 r/min, 0 HP	0 (0 W)	238
		1772 r/min, 1 HP	1 (735 W)	472
		1750 r/min, 2 HP	2 (1470 W)	473
		1730 r/min, 3 HP	3 (2205 W)	474
2	Inner race fault	1797 r/min, 0 HP	0 (0 W)	118
		1772 r/min, 1 HP	1 (735 W)	119
		1750 r/min, 2 HP	2 (1470 W)	119
		1730 r/min, 3 HP	3 (2205 W)	120
3	Outer race fault	1797 r/min, 0 HP	0 (0 W)	119
		1772 r/min, 1 HP	1 (735 W)	119
		1750 r/min, 2 HP	2 (1470 W)	118
		1730 r/min, 3 HP	3 (2205 W)	119
4	Rolling element fault	1797 r/min, 0 HP	0 (0 W)	119
		1772 r/min, 1 HP	1 (735 W)	118
		1750 r/min, 2 HP	2 (1470 W)	118
		1730 r/min, 3 HP	3 (2205 W)	118

5.2. Fault Diagnosis Based on VGF-GRSSL

The VGFs were extracted, as showed in the Figure 7. It is observed that the locations of features are separated according to their fault modes, remarkably. Therefore, the VGF method not only can extract fault features, but also has strong robustness to variable conditions.

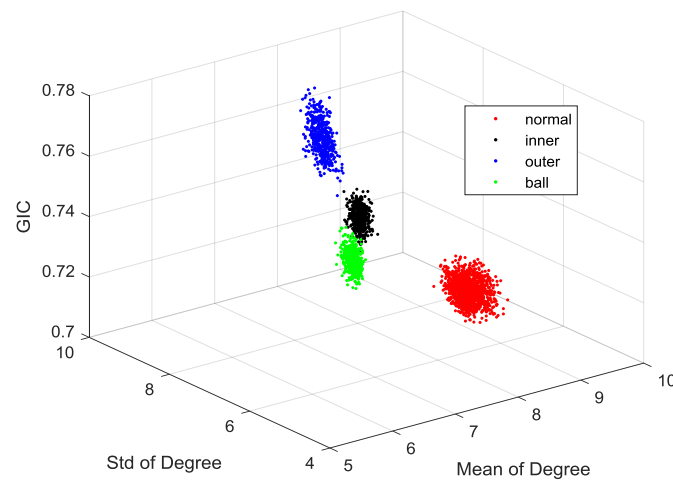


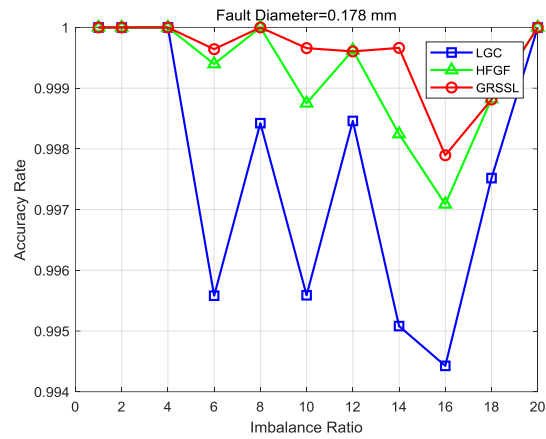
Figure 7. Feature space.

Based on the obtained features, the GRSSL method is employed for fault classification. For comparisons, GFHF and LGC are involved in this study. This paper set the hyper-parameter $\mu = 0.01$ for LGC and GRSSL. The algorithms GRSSL, GFHF, and LGC shared the same graph construction procedure. The standard k-nearest-neighbors (set $k = 6$) method was employed for the sparsification and Gaussian kernel weighting for reweighting the edges. In the Gaussian kernel, this study used the ℓ_2 distance, and the kernel bandwidth σ is defined as the average distance between each selected sample and its k'th nearest neighbor.

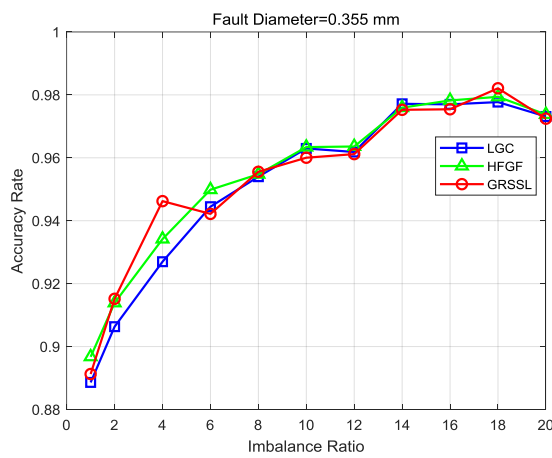
In the case of the multi-class problem, the imbalanced ratio is defined as the number of majority samples divided by the sum of number of minority samples. It is assumed that the number of each minority class contains the same number of samples.

To demonstrate the performance of the proposed method under different imbalanced ratios, for the labeled data, we fix the minority classes (including inner fault, outer fault, rolling element fault) to have only three label samples and then randomly select m labels from the majority class (normal). For the unlabeled data, we keep it with the same imbalanced ratio as the labeled data. Here, we varied the ratio ($r = m/3$) from 1 to 20.

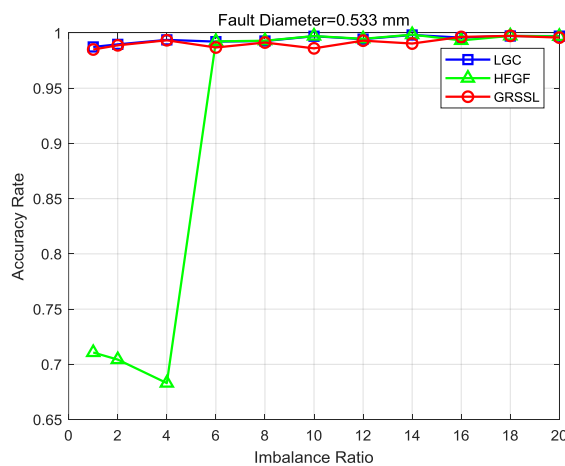
The results are shown in Figure 8 under three conditions (i.e., fault diameter 0.007, 0.014, and 0.021 inches (0.178 mm, 0.355 mm, and 0.533 mm)). In Figure 8a, it can be observed: (1) all three algorithms can obtain accuracy rates of nearly 100% when the imbalance ratio is less than 4, and (2) the accuracy of the GRSSL is more stable than the LGC and GFHF when the fault diameter is 0.007 inches (0.178 mm). Figure 8b illustrates results with a fault diameter of 0.014 inches (0.355 mm). The three methods obtained similar accuracy rates with varying imbalance ratios. Meanwhile, it is shown in Figure 8c that all methods achieved excellent accuracy rates except the GFHF with the imbalance ratio lower than 4. In summary, the GRSSL achieves a more stable accuracy rate than the LGC and GFHF under variable conditions with imbalanced unlabeled data.



(a)



(b)



(c)

Figure 8. Performance of Local and Global Consistency (LGC), Gaussian Fields and Harmonic Functions (GFHF), and graph-based rebalance semi-supervised learning (GRSSL) algorithms using the bearing data with a fault diameter of (a) 0.178 mm, (b) 0.355 mm and (c) 0.533 mm.

To illustrate the diagnostic performance of the proposed method exactly, we calculated the accuracy of each class, which is displayed in the Table 3. It is evident that the GRSSL method outperformed other methods under condition 1 and condition 3. However, all methods got a bad

performance for the inner race fault and ball fault under the condition 2. But they recognized the normal samples and faulty samples at rate of 100%, which is of significance to industrial field application.

Table 3. Results of LGC, GFHF, and GRSSL under different severity of fault with the imbalance ratio $r = 16$ (IRF-inner race fault, ORF-outer race fault, BF-ball Fault).

Fault Diameter	Method	Classification Accuracy (%)				Average Accuracy (%)
		Normal	IRF	ORF	BF	
Condition 1 (0.178 mm)	GRSSL	100.00	100.00	99.06	96.92	99.79
	LGC	100.00	94.46	97.84	97.06	99.44
	GFHF	100.00	97.72	99.04	97.68	99.71
Condition 2 (0.355 mm)	GRSSL	100.00	77.84	96.82	62.94	96.73
	LGC	100.00	84.56	98.54	71.24	97.61
	GFHF	100.00	82.64	98.78	71.92	97.56
Condition 3 (0.533 mm)	GRSSL	100.00	100.00	98.06	95.12	99.64
	LGC	100.00	97.48	96.04	96.08	99.46
	GFHF	100.00	98.68	97.06	95.68	99.55

6. Conclusions

Rolling element bearings are vital to rotating machineries. This paper proposes a novel hybrid method combined with visibility graph feature (VGF) and graph-based rebalance semi-supervised learning (GRSSL) for bearing fault diagnosis under variable conditions with imbalanced unlabeled data. Firstly, the visibility graph algorithm is used to extract features which are natively insensitive to variable conditions. Secondly, the GRSSL is utilized to make effective use of the imbalanced unlabeled data for fault classification. Experiment results have demonstrated the superiority of the proposed VGF and GRSSL: (1) Compared with approximate entropy, sample entropy, and fuzzy entropy, the VGF can effectively extract the fault characteristics under variable conditions; (2) GRSSL has superior performance in bearing fault diagnosis under variable conditions with imbalanced unlabeled data.

However, to some extent, the imbalanced ratio involved in this study only considered normal samples and faulty samples. The absence of faulty samples, such as inner race fault and outer race fault, may create new problems. Therefore, additional experiments under more different ratios should be done to validate and improve the method. Meanwhile, more attention should be paid to the development of semi-supervised learning diagnosis methods.

Author Contributions: X.C. collected and analyzed the data, made charts and diagrams, conceived and performed the experiments and wrote the paper; Z.W. and L.J. conceived the structure, provided guidance and modified the manuscript; Z.Z. analyzed the data and contributed analysis tools. Y.Q. provided guidance.

Funding: This research was funded by National Key R&D Program of China grant number 2016YFB1200402, the Fundamental Research Funds for the Central Universities (No. 2017RC011), the State Key Laboratory of Rail Traffic Control and Safety (Contract Nos. RCS2016ZQ003 and RCS2016ZT018).

Acknowledgments: This research is supported by National Engineering Laboratory for System Safety and Operation Assurance of Urban Rail Transit.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. IEEE Motor Reliability Working Group. Report of Large Motor Reliability Survey of Industrial and Commercial Installations. *IEEE Trans. Ind. Appl.* **1985**, *IA-4*, 853–872.
2. Henao, H.; Capolino, G.-A.; Fernandez-Cabanias, M.; Filippetti, F.; Bruzzese, C.; Strangas, E.; Pusca, R.; Estima, J.; Riera-Guasp, M.; Hedayati-Kia, S. Trends in Fault Diagnosis for Electrical Machines: A Review of Diagnostic Techniques. *IEEE Ind. Electron. Mag.* **2014**, *8*, 31–42. [[CrossRef](#)]
3. Yang, Y.; Yu, D.; Cheng, J. A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM. *Meas. J. Int. Meas. Confed.* **2007**, *40*, 943–950. [[CrossRef](#)]

4. Lou, X.; Loparo, K.A. Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mech. Syst. Signal Process.* **2004**, *18*, 1077–1095. [[CrossRef](#)]
5. Yu, D.; Cheng, J.; Yang, Y. Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings. *Mech. Syst. Signal Process.* **2005**, *19*, 259–270. [[CrossRef](#)]
6. Zheng, J.; Cheng, J.; Yang, Y. A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy. *Mech. Mach. Theory* **2013**, *70*, 441–453. [[CrossRef](#)]
7. Shuang, L. Bearing Fault Diagnosis Based on PCA and SVM. In Proceedings of the International Conference on Mechatronics and Automation, Harbin, China, 5–8 August 2007; pp. 3503–3507. [[CrossRef](#)]
8. Sugumaran, V.; Muralidharan, V.; Ramachandran, K.I. Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing. *Mech. Syst. Signal Process.* **2007**, *21*, 930–942. [[CrossRef](#)]
9. Li, B.; Chow, M.Y.; Tipsuwan, Y.; Hung, J.C. Neural-network-based motor rolling bearing fault diagnosis. *IEEE Trans. Ind. Electron.* **2000**, *47*, 1060–1069. [[CrossRef](#)]
10. Immovilli, F.; Bellini, A.; Rubini, R.; Tassoni, C. Diagnosis of bearing faults in induction machines by vibration or current signals: A critical comparison. *IEEE Trans. Ind. Appl.* **2010**, *46*, 1350–1359. [[CrossRef](#)]
11. Frosini, L.; Harlisca, C.; Szabo, L. Induction machine bearing fault detection by means of statistical processing of the stray flux measurement. *IEEE Trans. Ind. Electron.* **2015**, *62*, 1846–1854. [[CrossRef](#)]
12. Jiang, L.; Xuan, J.; Shi, T. Feature extraction based on semi-supervised kernel Marginal Fisher analysis and its application in bearing fault diagnosis. *Mech. Syst. Signal Process.* **2013**, *41*, 113–126. [[CrossRef](#)]
13. Hu, Y.; Baraldi, P.; Di Maio, F.; Zio, E. A Systematic Semi-Supervised Self-adaptable Fault Diagnostics approach in an evolving environment. *Mech. Syst. Signal Process.* **2017**, *88*, 413–427. [[CrossRef](#)]
14. Zhao, X.; Li, M.; Xu, J.; Song, G. An effective procedure exploiting unlabeled data to build monitoring system. *Expert Syst. Appl.* **2011**, *38*, 10199–10204. [[CrossRef](#)]
15. LIU, T.; LI, G. The imbalanced data problem in the fault diagnosis of rolling bearing. *Comput. Eng. Sci.* **2010**, *5*, 44.
16. Wang, J.; Jebara, T.; Chang, S.F. Semi-Supervised Learning Using Greedy Max-Cut. *J. Mach. Learn. Res.* **2013**, *14*, 771–800.
17. Zhu, X.; Ghahramani, Z.; Lafferty, J.D. Semi-supervised learning using gaussian fields and harmonic functions. In Proceedings of the 20th International Conference on Machine Learning (ICML-03), Washington, DC, USA, 21–24 August 2003; pp. 912–919.
18. Zhou, D.; Bousquet, O.; Lal, T.N.; Weston, J.; Schölkopf, B. Learning with Local and Global Consistency. *Adv. Neural Inf. Process. Syst.* **2004**, *1*, 321–328.
19. Borghesani, P.; Ricci, R.; Chatterton, S.; Pennacchi, P. A new procedure for using envelope analysis for rolling element bearing diagnostics in variable operating conditions. *Mech. Syst. Signal Process.* **2013**, *38*, 23–35. [[CrossRef](#)]
20. Feng, Z.; Chen, X.; Wang, T. Time-varying demodulation analysis for rolling bearing fault diagnosis under variable speed conditions. *J. Sound Vib.* **2017**, *400*, 71–85. [[CrossRef](#)]
21. Tian, Y.; Ma, J.; Lu, C.; Wang, Z. Rolling bearing fault diagnosis under variable conditions using LMD-SVD and extreme learning machine. *Mech. Mach. Theory* **2015**, *90*, 175–186. [[CrossRef](#)]
22. Wang, Z.; Jia, L.; Qin, Y. Adaptive Diagnosis for Rotating Machineries Using Information Geometrical Kernel-ELM Based on VMD-SVD. *Entropy* **2018**, *20*, 73. [[CrossRef](#)]
23. Lacasa, L.; Luque, B.; Ballesteros, F.; Luque, J.; Nuno, J.C. From time series to complex networks: The visibility graph. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4972–4975. [[CrossRef](#)] [[PubMed](#)]
24. Chapelle, O.; Scholkopf, B.; Zien, A. Semi-Supervised Learning [Book Reviews]. *IEEE Trans. Neural Netw.* **2009**, *20*, 542. [[CrossRef](#)]
25. Li, C.; Zhou, J. Semi-supervised weighted kernel clustering based on gravitational search for fault diagnosis. *ISA Trans.* **2014**, *53*, 1534–1543. [[CrossRef](#)] [[PubMed](#)]
26. Qin, W.L.; Zhang, W.J.; Wang, Z.Y. Improvement of Roller Bearing Diagnosis with Unlabeled Data Using Cut Edge Weight Confidence Based Tritraining. *Shock Vib.* **2016**, *2016*, 1646898. [[CrossRef](#)]
27. Zhao, M.; Li, B.; Qi, J.; Ding, Y. Semi-supervised classification for rolling fault diagnosis via robust sparse and low-rank model. In Proceedings of the IEEE 15th International Conference on Industrial Informatics (INDIN), Emden, Germany, 24–26 July 2017; pp. 1062–1067. [[CrossRef](#)]

28. Subramanya, A.; Bilmes, J. Semi-Supervised Learning with Measure Propagation. *J. Mach. Learn. Res.* **2011**, *12*, 3311–3370.
29. Wang, C.-C.; Too, G.-P.J. Rotating machine fault detection based on HOS and artificial neural networks. *J. Intell. Manuf.* **2002**, *13*, 283–293. [[CrossRef](#)]
30. Kim, J.; Wilhelm, T. What is a complex graph? *Phys. A Stat. Mech. Appl.* **2008**, *387*, 2637–2652. [[CrossRef](#)]
31. Chung, F.R.K. *Spectral Graph Theory*; American Mathematical Society: Providence, RI, USA, 1997; ISBN 0821803158.
32. Karp, R.M. Reducibility among combinatorial problems. In *Complexity of Computer Computations*; Springer: Berlin, Germany, 1972; pp. 85–103.
33. Tse, P.W.; Wang, D. The design of a new sparsogram for fast bearing fault diagnosis: Part 1 of the two related manuscripts that have a joint title as “two automatic vibration-based fault diagnostic methods using the novel sparsity measurement—Parts 1 and 2”. *Mech. Syst. Signal Process.* **2013**, *40*, 499–519. [[CrossRef](#)]
34. Yan, R.; Gao, R.X. Approximate entropy as a diagnostic tool for machine health monitoring. *Mech. Syst. Signal Process.* **2007**, *21*, 824–839. [[CrossRef](#)]
35. Han, M.; Pan, J. A fault diagnosis method combined with LMD, sample entropy and energy ratio for roller bearings. *Measurement* **2015**, *76*, 7–19. [[CrossRef](#)]
36. Li, Y.; Xu, M.; Wang, R.; Huang, W. A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy. *J. Sound Vib.* **2016**, *360*, 277–299. [[CrossRef](#)]
37. Case Western Reserve University. Bearing Data Center. Available online: <http://csegroups.case.edu/bearingdatacenter/home> (accessed on 18 January 2018).



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).