We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter 6

Fault Diagnosis and Health Assessment for Rotating Machinery Based on Kernel Density Estimation and Kullback-Leibler Divergence

Yu Liu, Chen-Yao Yan and Fan Zhang

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/67360

Abstract

To avoid severe damages and unexpected shutdowns, fault diagnosis and health assessment of rotating machinery have received considerable attention in recent years. On the other hand, as a great amount of data become acquirable and accessible in industry, data-driven tools have become an emerging research area, acting as a complement to the model-based (or physics-based) fault diagnosis and health assessment methods. In this chapter, based on the kernel density estimation (KDE) and the Kullback-Leibler divergence (KLID), a new data-driven fault diagnosis approach and a new health assessment approach are introduced. By utilizing the KDE, the statistical distribution of selected features can be readily estimated without assuming any parametric family of distributions, whereas the KLID is able to quantify the discrepancy between two probability distributions of selected features. An integrated Kullback-Leibler divergence, which aggregates the KLID of all the selected features, is introduced to discriminate various fault types or health status of rotating machinery. The effectiveness of the proposed approaches is demonstrated through three case studies of fault diagnosis and health assessment of rotating machinery.

Keywords: data-driven approach, fault diagnosis, health assessment, kernel density estimation, kullback-Leibler divergence, rotating machinery

1. Introduction

Rotating machinery has widespread applications in advanced manufacturing and engineering systems, e.g., wind turbines, power generators, and machining tools. The crucial components in rotating machinery, such as bearings and gears, are oftentimes suffering undesirable

INTECH
open science | open minds

© 2017 The Author(s). Licensee InTech. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

stresses and sudden shocks under which initial defects will appear [1]. If maintenance activities cannot be taken properly and timely, tiny defects will gradually propagate and eventually cause severe damages and unexpected shutdown to the entire systems. It is, therefore, of paramount importance to accurately detect the presence of faults as early as possible and track the growth of the tiny faults to avoid the consequence of severe damages caused by faults and also facilitate preventive maintenance planning before the complete failure of engineering systems [1].

Fault diagnosis and health assessment are the two important tools for detecting the operating condition of rotating machinery based on which preventative maintenance can be scheduled in a timely manner. In general, existing methods for fault diagnosis and health assessment can be classified into two categories [2]: model-based (or physics-based) approaches and data-driven approaches. The model-based approaches require specific mechanistic knowledge and theory relevant to the monitored machine, and a particular fault or health status of a system can be determined by comparing available system measurements with a priori information represented by the corresponding system's analytical/mathematical model [3]. These methods could be very accurate when a correct model can be built up. For example, several models have been developed to characterize the crack growth [4, 5]. However, due to the limited knowledge of underlying mechanisms and physics, it becomes very difficult, or even impossible, to fully understand the evolution of defects and faults of complex engineered systems [1, 2]. With the fast development of condition monitoring and intelligent computing technologies, data-driven approaches have received considerable attention in recent years. Many advanced classification methods have been applied to data-driven fault diagnosis [6–9]. Among them, support vector machine [9, 10] and artificial neural network (ANN) [11] are two representative and powerful classification methods, and they have been extensively used in fault diagnosis for rotating machinery [9–13]. By using the data-driven approaches, the fault type or health status of a system can be mapped into the feature space [1, 2]. In other words, the relation between features extracted from condition monitoring data and fault model/damage levels can be acquired from a set of training data. Thereby, compared to the model-based approaches, the data-driven approaches possess two merits: (1) by the data-driven approaches, fault diagnosis or health assessment can be executed automatically without heavy involvement of engineers and (2) unlike the model-based approaches that need professional expertise to make judgments, the data-driven approaches do not rely on expertise and knowledge from experts too much [13].

In most cases, a data-driven approach for fault diagnosis and health assessment of rotating machinery consists of five basic steps as shown in **Figure 1**. The raw data, e.g., vibration signals, collected from condition monitoring serve as inputs of a data-driven fault diagnosis or health assessment approach. Subsequently, by using advanced signal processing algorithms, e.g., the fast Fourier transform (FFT), the empirical model decomposition (EMD), and the wavelet transform, a bunch of features which are more or less relevant to the health status of the monitored device can be extracted from the raw data. A subset of the most significant features, which are sensitive to a specific fault type or health status of the system, will be chosen from all the extracted features. On the other hand, irrelevant and redundant features can be eliminated at this stage to mitigate the computational burden and improve the accuracy

http://dx.doi.org/10.5772/67360

of results. It is followed by the fault classification or health assessment where the selected features will be used as inputs of fault/health status classifier. Many advanced classification methods can be applied. Among many, the support vector machine (SVM) [10] and the artificial neural network (ANN) [11] are two representative approaches.

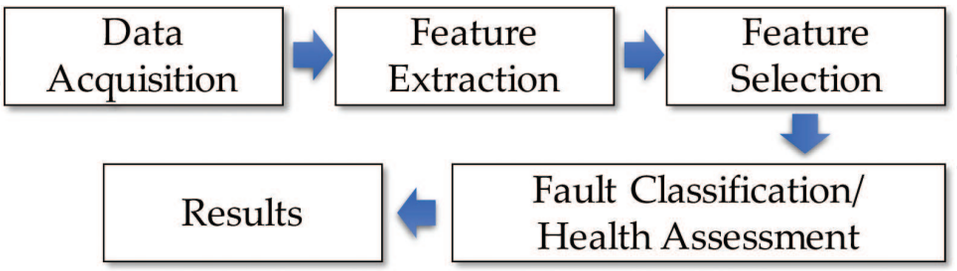


Figure 1. The basic procedures of data-driven fault diagnosis or health assessment methods.

In this chapter, a new data-driven fault diagnosis approach and a data-driven health assessment approach are put forth. Two statistical tools, i.e., the kernel density estimation (KDE) and the Kullback-Leibler divergence (KLID), are used jointly to identify fault modes/health status of rotating machinery from a statistical viewpoint. The KDE, which is a nonparametric probability density estimation approach, is able to adaptively fit a data set to a smooth density function without pre-specifying a particular distribution type [14, 15]. On the other hand, the KLID, which is so-called information divergence or relative entropy, is a measure of the discrepancy between two probability distributions [16]. By using the KDE and KLID jointly, an integrated Kullback-Leibler divergence can be developed to identify faults modes/health status of rotating machinery.

The rest of this chapter is rolled out as follows: Section 2 introduces the principle of the KDE and the KLID. The proposed fault diagnosis approach together with two case studies is presented in Section 3. In Section 4, the proposed health assessment approach and its application to a case study are elaborated. The chapter closes with a brief conclusion in Section 5.

2. The kernel density estimation (KDE) and the Kullback-Leibler divergence (KLID)

2.1. The kernel density estimation

The kernel density estimation originally introduced by Rosenblatt and Parzen [14, 15] is a nonparametric tool to infer the probability density function of a data set. It stems from the empirical probability density function.

Let X_1, X_2, ⋯, X_n represent n independent and identically distributed (i.i.d.) random samples from a random quantity X with an unknown probability density function $f(x)$. The kernel density function is defined as:

$$\widehat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) \tag{1}$$

where $K(\cdot)$, a symmetric function with integration equal to 1, is the kernel function. The kernel function may not be necessary a position function but has to guarantee $\widehat{f}_h(x)$ satisfies the basic requirement of a probability density function. Many different types of kernel functions have been proposed in Ref. [15], e.g., uniform, Gaussian, triangle, Epanechnikov, and quaritic. Particularly, the Gaussian kernel function, which has been extensively adopted due to many mathematical properties, such as centrality and gradual decay, is formulated as:

$$K(u) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^2\right) \tag{2}$$

The bandwidth h ($h > 0$) of the kernel function has a heavy influence on the smoothness of $\widehat{f}_h(x)$. A larger h indicates that a greater region of samples around the centre point x influences the probability density estimation, vice versa. A proper setting for the bandwidth h is, therefore, of great significance for the KDE. The mean integrated squared error (MISE) is the most common optimality criterion to choose a proper bandwidth, and it is defined as [15]:

$$MISE(h) = E\int\left(\widehat{f}_h(x) - f(x)\right)^2 dx \tag{3}$$

Under weak assumptions on $f(\cdot)$ and $K(\cdot)$, one has:

$$MISE(h) = AMISE(h) + o\left(\frac{1}{nh} + h^4\right) \tag{4}$$

where $o(\cdot)$ is infinitesimal. The AMISE is the asymptotic MISE, and it is defined as:

$$AMISE(h) = \frac{R(K(\cdot))}{nh} + \frac{1}{4}m_2(K(\cdot))^2 h^4 R\left(f''(\cdot)\right) \tag{5}$$

where $R(g) = \int g(x)^2 dx$; $m_2(K) = \int x^2 K(x) dx$; $f''(\cdot)$ is the second-order derivative of $f(\cdot)$; and n is the total number of samples. The following differential equation can be used to seek the minimal value of the AMISE as:

$$\frac{\partial}{\partial h}AMISE(h) = -\frac{R(K(\cdot))}{nh^2} + m_2(K(\cdot))^2 h^3 R\left(f''(\cdot)\right) = 0 \tag{6}$$

Thus, the minimal value of h is:

$$h^*_{AMISE} = \frac{R(K(\cdot))^{1/5}}{m_2(K(\cdot))^{2/5} R\left(f''(\cdot)\right)^{1/5} n^{1/5}} \tag{7}$$

It should be noted that the above equation is implicit and contains the unknown density function $f(\cdot)$ or $f''(\cdot)$. In engineering practice, the density to be estimated is also Gaussian if the Gaussian basis function is used to approximate univariate data. Thus, the optimal value of h is:

http://dx.doi.org/10.5772/67360

$$h^*_{AMISE} = \left(\frac{4\widehat{\sigma}^5}{3n}\right)^{\frac{1}{5}} \approx 1.06\widehat{\sigma}n^{-\frac{1}{5}} \tag{8}$$

where $\widehat{\sigma}$ is the standard deviation of samples. Such approximation called the Gaussian approximation is adopted in this work.

2.2. The Kullback-Leibler divergence

The Kullback-Leibler divergence (KLID) was first introduced by Solomon Kullback and Richard Leibler in 1951 [16], and it has been applied to quantify the difference of two distributions. For two discrete probability distributions P and Q, the KLID of Q from P is written as:

$$D_{KL}(P||Q) = \sum_i \ln\left(\frac{P(i)}{Q(i)}\right)P(i) \tag{9}$$

In essence, Eq. (9) is the expectation of the logarithmic difference between the probabilities P and Q, and the expectation is taken by the probability P. The KLID is valid if the integration of P and Q is both equal to 1. If $Q(i) = 0$, then $P(i) = 0$ for all i. For the case where $P(i) = 0$ and $P(i)/Q(i) = 0$, $\ln(P(i)/Q(i))P(i) = 0$ since $\lim_{x\to 0} x\ln(x) = 0$.

Based on the Gibbs' inequality, $D_{KL}(P||Q) = 0$ if and only if $P = Q$ holds almost everywhere. A smaller value of $D_{KL}(P||Q)$ represents a greater similarity between the two probability distributions. It is noteworthy that although the KLID can quantify the distance between two probability distributions, it does not fully satisfy some important properties of distance measure, e.g., symmetry and triangle inequality. For instance, the KLID of P over Q is generally not exactly equal to the KLID of Q over P. Nevertheless, the symmetry property is very crucial in the classification issue. In our work, the symmetrized distance of KLID defined in Ref. [12] is adopted to measure the discrepancy between two probability distributions, and it is formulated as:

$$D_{KL}(P,Q) = \frac{1}{2}[D_{KL}(P||Q) + D_{KL}(Q||P)]. \tag{10}$$

3. The proposed fault diagnosis approach

Follow the basic procedures of data-driven fault diagnosis method as shown in **Figure 1**, the proposed approach for the fault diagnosis of rotating machinery is given in **Figure 2**. A set of time- and frequency-domain features will be first extracted from the raw vibration signals by the ensemble empirical mode decomposition (EEMD), the Hilbert Transform, and so on. The distance-based feature selection method will be used to identify a subset of sensitive features. The kernel density estimation (KDE) and the Kullback-Leibler divergence (KLID) introduced in Section 2 will be used together as a new classifier to discriminate various fault types.



Figure 2. The procedures of the proposed fault diagnosis approach for rotating machinery.

3.1. The details of the proposed method

In this section, feature extraction, feature selection, kernel density estimation, and Kullback-Leibler divergence will be integrated together to realize the fault diagnosis for rotating machinery. Some important symbols to be used hereinafter are explained here:

1. KD_i^j (j = 1, 2, ⋯, n; i = 1, 2, ⋯, C) denotes the KDE function of the jth feature of the training samples for type i fault. The vector $\mathbf{KD}_i = \left(KD_i^1, KD_i^2, \cdots, KD_i^n\right)$ is the KDE function set of all the n selected features of the training samples for type i fault.

2. TKD_i^j(j = 1, 2, ⋯, n;i = 1, 2, ⋯, C) is the KDE function of the jth feature of the training samples for type i fault after adding a testing sample. The vector $\mathbf{TKD}_i = \left(TKD_i^1, TKD_i^2, \cdots, TKD_i^n\right)$ is the KDE function set of all the n selected features of the training samples for type i fault after a testing sample is added.

3. KL_i^j (j = 1, 2, ⋯, n; i = 1, 2, ⋯, C) is the KLID between KD_i^j and TKD_i^j. The vector $\mathbf{KL}_i = \left(KL_i^1, KL_i^2, \cdots, KL_i^n\right)$ contains the KLIDs of all the n selected features.

The overall flowchart of the proposed approach for classifying two fault types is shown in **Figure 3**.

In **Figure 3**, the proposed method is illustrated through classifying two fault models, i.e., type I fault and type II fault. The sample sets from these two types of fault modes act as the training sample sets, whereas one sample set with unknown fault mode serves as the testing sample set to be classified. Nine time-domain features together with 10 frequency-domain features are extracted from the raw vibration signal and the first four IMFs decomposed by the EEMD. The technical details of the EEMD can be found in Refs. [17, 18]. Thus, the original feature set consists of 95 features. The distance-based evaluation approach is, then, applied to assess the effectiveness of each feature. The corresponding effectiveness factor of the jth (j = 1, 2, ⋯, 95) feature is denoted as α_j (see Ref. [19] for more details on the distance-based evaluation approach). The features with a larger effectiveness factor are more sensitive to these particular fault types. By sorting all the features by their effectiveness factors in a descending order, the first m features are selected from the original feature set and serve as the inputs of the classifier. Thereby, the importance of the ^{j}th feature to the fault classification is formulated as:

http://dx.doi.org/10.5772/67360

$$F_j = \frac{\alpha_j}{\sum_{i=1}^{n} \alpha_i}, (j = 1, 2, \cdots, m) \tag{11}$$

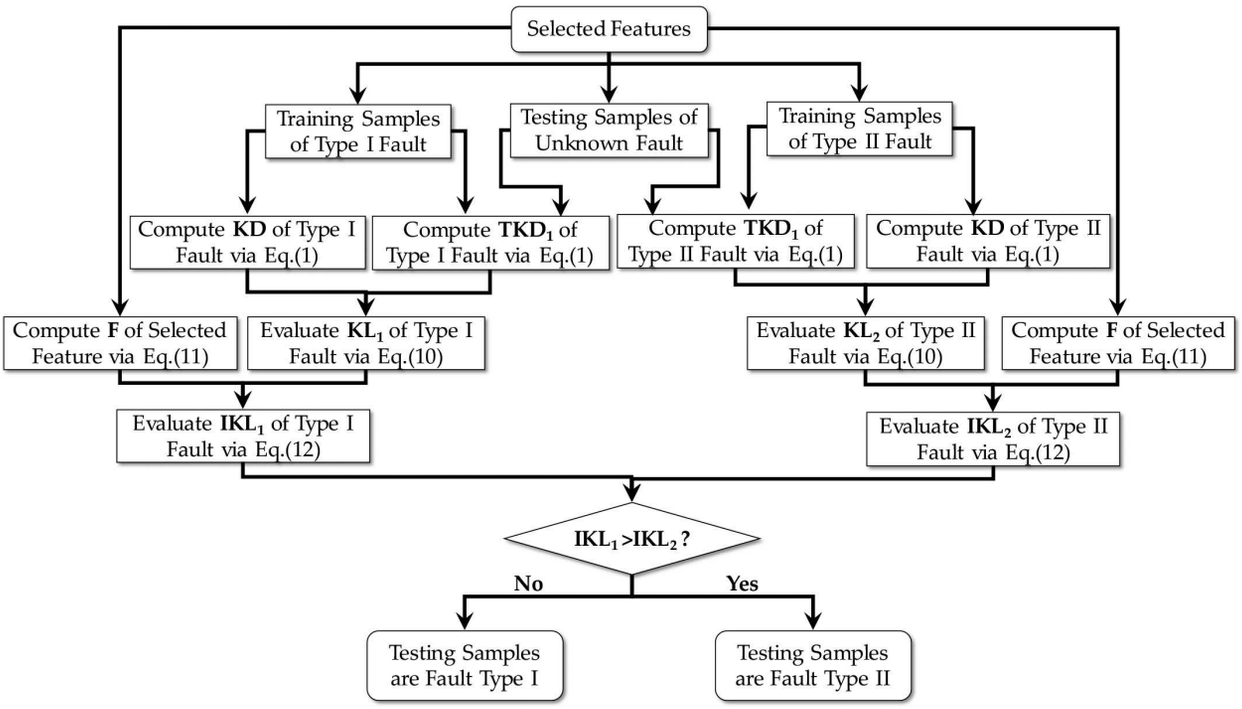


Figure 3. The flowchart of the proposed fault diagnosis approach for the case with two fault types.

The probability density of the selected features of each training set can be then characterized by the kernel density function. For instance, KD_1^1 and KD_2^1 are the first feature of type I and type II faults, respectively, are shown in **Figure 3**. If one sample from the testing sample set is added into the two training sets, respectively, and the corresponding probability distributions of the two new sample sets of the first feature can be also estimated by the kernel density function and denoted as TKD_1^1 and TKD_2^1, respectively. In the same manner, KD_1^j, KD_2^j, TKD_1^j, and TKD_2^j ($j = 1,2,3,....,m$) for all the selected features can be estimated. It is followed by computing the KL_1^j and KL_2^j, the symmetrized Kullback-Leibler divergences (KLIDs) between KD_1^j and TKD_1^j, KD_2^j and TKD_2^j ($j = 1, 2, \cdots, m$), via Eq. (10). To get an overall assessment of all the n selected features, an integrated KLID, denoted as IKL_i, is defined here to aggregate all the symmetrized KLIDs KL_i^j for the type i fault together as:

$$IKL_i = \sum_{j=1}^{m} F_i \times KL_i^j \tag{12}$$

where $F_j (j = 1, 2, \cdots, m)$ computed by Eq. (11) is the importance of the jth feature and $\mathbf{F} = (F_1, F_2, \cdots, F_m)$. Using Eq. (12), IKL_1 and IKL_2 of any sample from the testing sample set

with respect to the type I and type II faults can be obtained. A smaller value of IKL_i implies that the testing sample is similar to the corresponding training sample set in a statistic sense. Put another way, adding the testing sample into the training sample set causes a slight influence on the statistical distribution of the training sample set. Hence, the fault type of the testing sample can be discriminated. For instance, if $IKL_1 > IKL_2$, it can be concluded that the fault implied by the testing sample is more prone to the type II fault than type I fault. By this way, all the testing sample sets can be classified into one of the two fault types.

Following the same manner, the proposed method can be straightforwardly applied to a more general case in which the number of fault modes/damage levels to be classified is greater than two. In this case, the fault modes/damage level of the testing sample can be distinguished by looking for the smallest integrated KLID among all the known fault modes/damage levels.

3.2. Two case studies

The effectiveness of the proposed method in terms of diagnosing rotating machinery faults is validated in this section through two case studies of the bevel gears and the rolling element bearings.

3.2.1. Experimental rigs

Case 1: Experiments are performed on a machinery fault simulator produced by Spectra Quest, Inc. The experimental setup and the bevel gears to be tested are presented in **Figure 4**. The experimental setup composed of a motor, a coupling, bearings, two bevel gearboxes (one good right angle gearbox and one worn right angle gearbox), discs, belts, and a shaft. The bevel gearbox is driven by an AC motor and coupled with rub belts. The rotation speed was fixed to 1800 r/min. Three faulty gears, i.e., worn gear, gear with missing teeth, and gear with broken tooth, were simulated on the experimental setup. The raw vibration data were collected by an accelerometer that was mounted on the top of the gearbox. The data sampling rate was 20 kHz, and the data length was 4096 points [20].



Figure 4. The experiment rig and the four bevel gears with different damages. (a) Normal gear, (b) gear with broken tooth, (c) gear with missing teeth, and (d) gear with worn tooth.

http://dx.doi.org/10.5772/67360

Case 2: The experimental data are from the Case Western Reserve University [21]. The experimental rig is consisting of the Reliance Electric 2HP IQPreAlet, which is connected to a dynamometer. The bearings supporting the motor shaft were examined. Faults were artificially generated by creating crack size of 0.007, 0.014, 0.021, and 0.028 inches on the drive-end bearing through the electric discharge machining. These faults are separately distributed on the inner raceway, rolling element, and outer raceway. The raw vibration signals were collected by the two accelerometers mounted on the motor housing and the outer race of the drive-end bearing. The sampling frequency was set to be 12 kHz, whereas the sampling length was 12 k. The rotating speed was 1750 r/min. The detailed settings of this experiment can be found in Ref. [21].

3.2.2. Experimental testing and results

The raw data from the above two experimental setups are used to validate the proposed method. Without loss of generality, the data sets with the same type of defects or severity are randomly divided into training samples and testing samples. **Table 1** gives the training and testing sample sizes, the places of defects, and the defect sizes of the two case studies. The data set A in **Table 1** is from the Case 1, whereas the data sets B and C come from the Case 2. The capability of the proposed method in distinguishing the types of defects is examined through the data sets A and B. The capability of the method in identifying the severity of the same type of defect is validated through the data set C.

Data set		Number of training samples	Number of testing samples	Defect size (inch) (training/testing)*	Condition
A		35	35	–	Normal
		35	35	–	Broken tooth
		35	35	–	Missing teeth
		35	35	–	Worn tooth
B	B_1	35	35	0.007/0.021	Inner race
		35	35	0.007/0.021	Ball
	B_2	35	35	0.021/0.007	Inner race
		35	35	0.021/0.007	Ball
C		35	35	0.007	
		35	35	0.014	Inner race
		35	35	0.021	

*'-' for the data set A denotes the defect sizes of the training and testing samples are exactly the same but unmeasurable by a physical dimension.

Table 1. The data sets for defect and severity classification.

Two hundred and eighty data sets with four different operation conditions, i.e., normal condition, bevel gear with broken tooth, bevel gear with missing teeth, and bevel gear with worn tooth, are included in the data set A. The defect sizes of training and testing sample sets are exactly the same. It can be, thus, regarded as a four-class classification problem.

The data set B composed of 280 data sets of the faulty bearings has only two types of fault modes, i.e., inner race fault and ball fault. The data set B is divided into two subsets, i.e., subsets B1 and B2. Each of the subsets has 140 data samples. The study can be conducted to investigate the effectiveness of the proposed method if the fault mode of the training sample set is exactly the same as the testing sample set but the defect sizes are different. For the subset B1, 70 samples with the fault detect size of 0.007 inches are treated as the training sample set, whereas the remaining 70 samples with the fault detect size of 0.021 inches are the testing samples. To the contrary, for the subset B2, the training sample set of the subset B1 is treated as the testing set of the subset B2 and the testing set of the subset B1 is treated as the training sample set of the subset B2.

The data set C is consisting of 210 samples. The data set C is collected from the case where a defect is on the inner race. The defect sizes for the data set are 0.007, 0.021, and 0.028 inches. The aim of examining these three data sets is to validate the effectiveness of the proposed method in terms of identifying the defect severity (damage levels).

We exemplify the implementation of the proposed method to the data set A. Ninety-five time- and frequency-domain features are firstly extracted from the data set A. The effectiveness factors α_j of all the 95 features computed by the distance evaluation approach are shown in **Figure 5**, and the first 10 features with the greatest values are selected among all the 95 features.

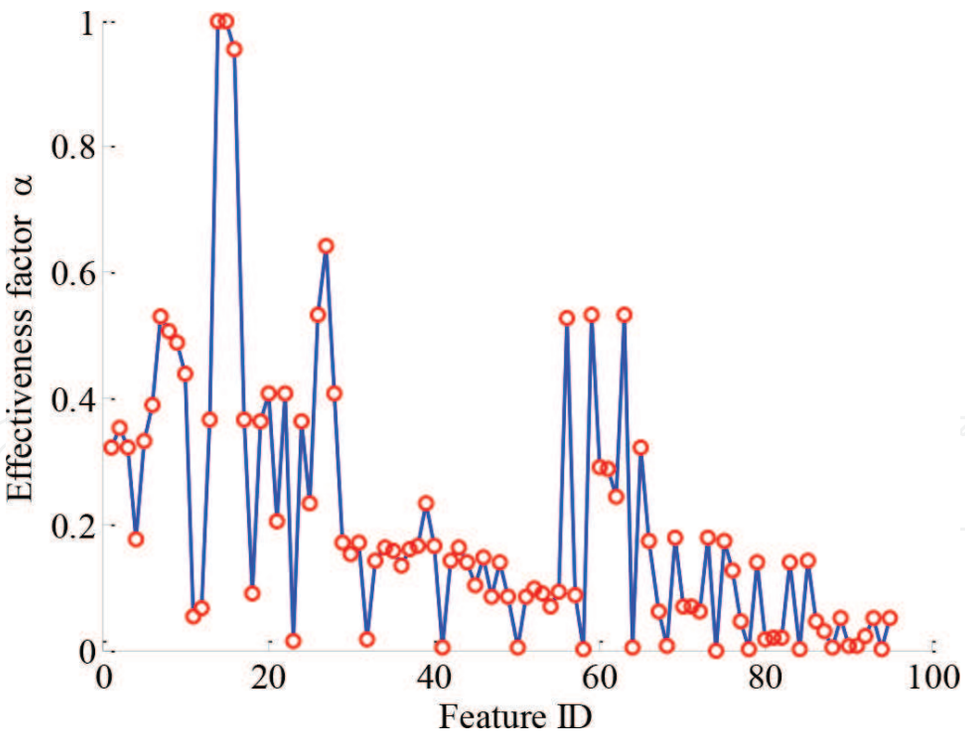


Figure 5. The effectiveness factor of all the extracted features.

Consequently, the probability density functions of the *j*th feature of the training sample sets for the four conditions, i.e., bevel gears with normal, broken tooth, missing teeth, and worn tooth

http://dx.doi.org/10.5772/67360

conditions respectively, can be obtained by the KDE, and they are denoted as $KD_i^j(i = 1,2,3,4)$. In the next step, a sample randomly picked up from the testing sample sets is included into the four training sets, and then, the corresponding probability density functions $TKD_i^j(i = 1,2,3,4)$ for the new sample set can be estimated. **Figures 6** and **7** give two examples of the probability density functions TKD_i^j of the first feature for the four training sample sets when a sample from the one of the four testing sampling sets is added.

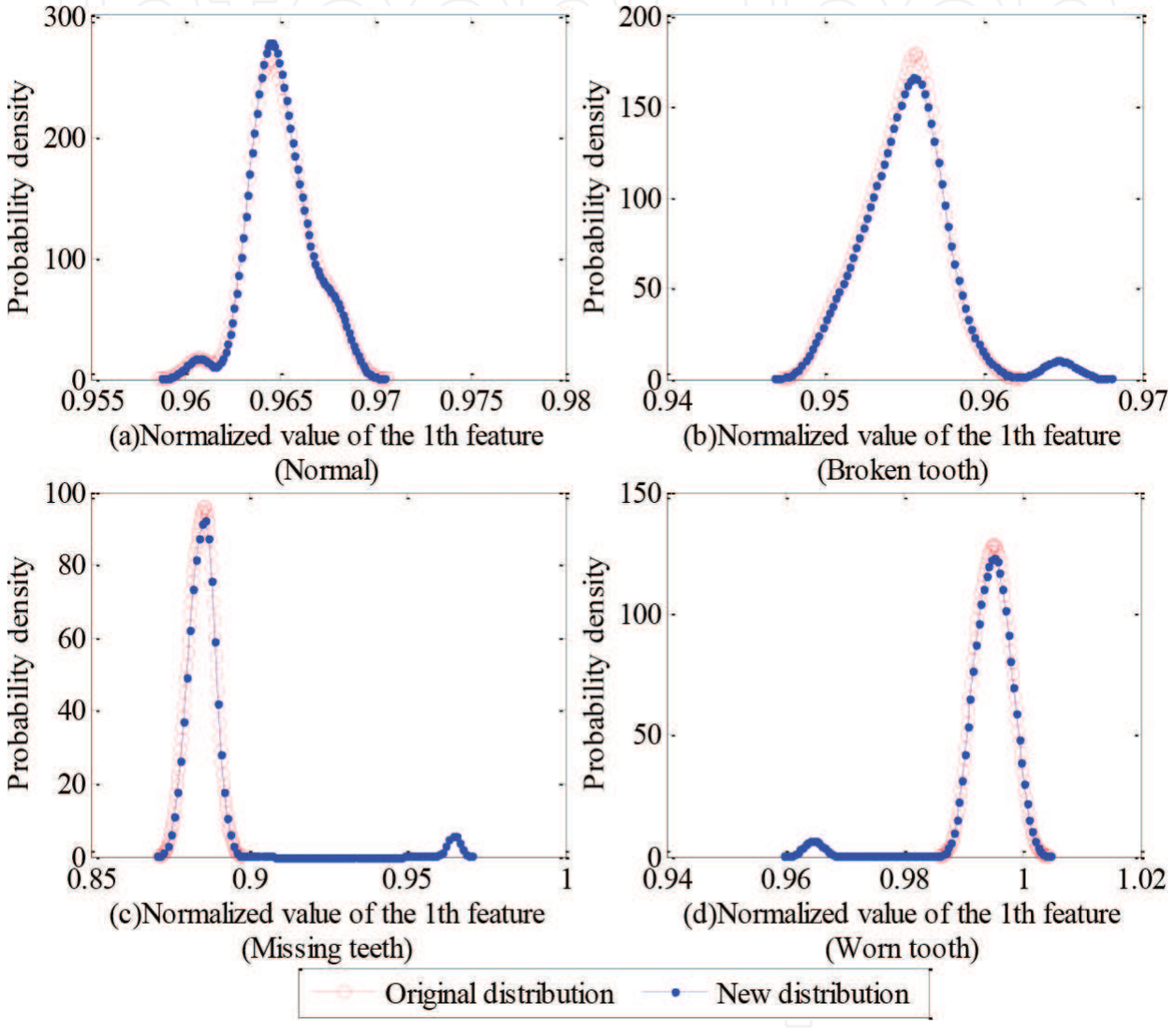


Figure 6. The original probability density functions of the four training sample sets and the corresponding new probability density functions after adding a normal testing sample.

In **Figures 6** and **7**, the red curves with circles are the original probability density functions of the first feature of the corresponding training set. The blue curves with dots represent the new probability density functions when a testing sample is included. For instance, as observed in **Figure 6(a)**, when a testing sample from the normal condition is added, the new probability density function of the first feature is almost the same as the original probability density function.

However, as seen from **Figure 6(b)–(d)**, the probability density functions exhibit a larger discrepancy when the testing sample from the normal condition is included to the other three training sample sets. Because the statistical characteristics of the first feature of the testing sample from the normal condition are distinct from these samples from the other three conditions, and the new probability density functions, therefore, generate a larger discrepancy from the original ones. Likewise, as observed in **Figure 7**, if the conditions of the new sample and the training sample sets (i.e., missing teeth) are the same, the new sample added to the training sample sets has minor impact on the probability density functions, otherwise a greater influence can be seen.

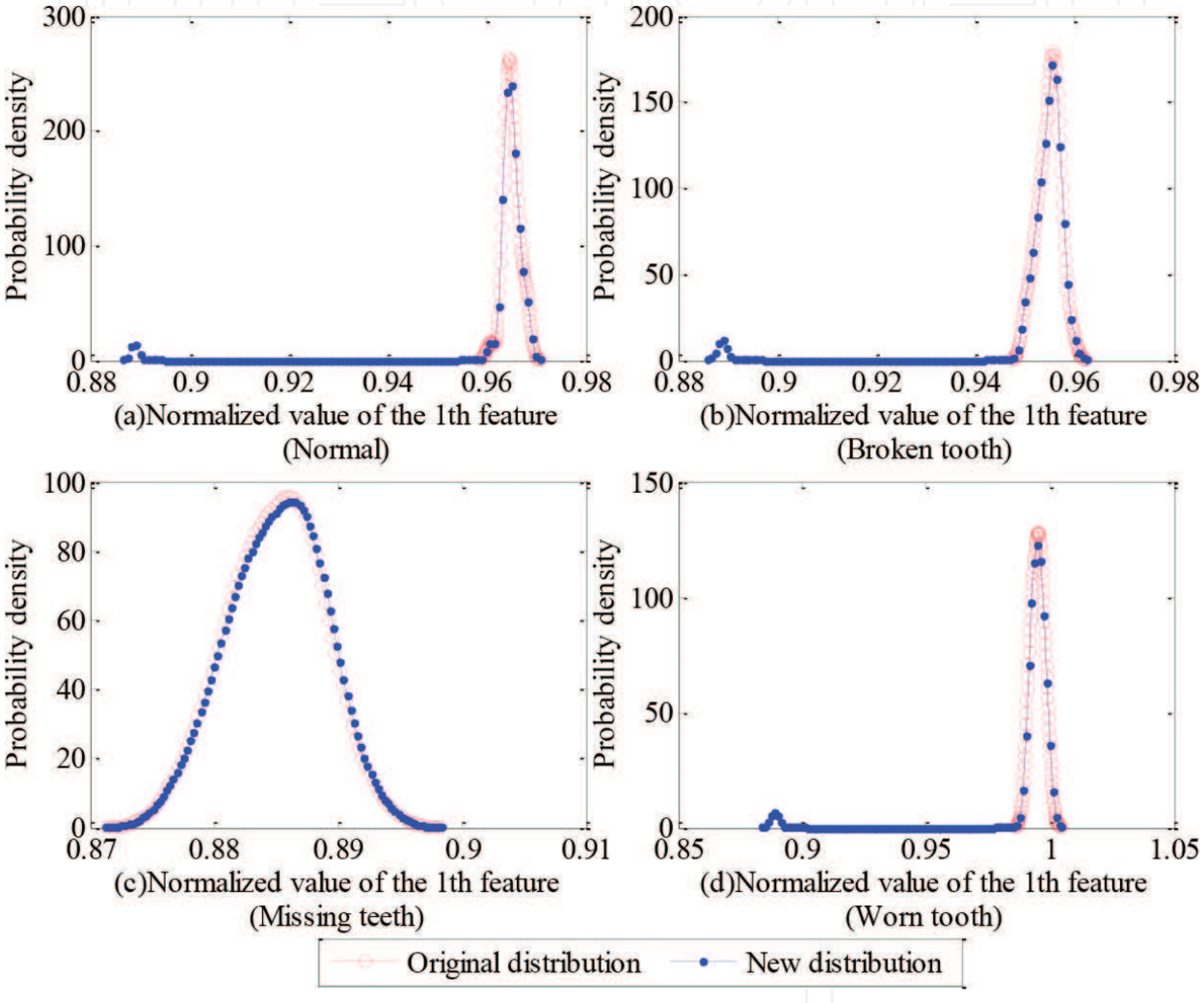


Figure 7. The original probability density functions of the four training sample sets and the corresponding probability density functions after adding a testing sample with missing teeth.

In the next step, the KLID is used to measure the difference between the original and the new distributions of the first feature in a quantitative way. The results are denoted as $KL_i^1 (i = 1,2,3,4)$ for the four different conditions. Following the same manner, the KLIDs can be evaluated for all the selected features. The integrated KLIDs, denoted as $IKL_i (i = 1,2,3,4)$, that aggregate the KLIDs of all the selected features are assessed based on the weights of the 10 selected features through Eq. (12).

http://dx.doi.org/10.5772/67360

The classification accuracy that is measured by evaluating the percentage of correctly distinguishing the fault modes or defect levels for the three data sets is presented in **Table 2**. A greater value of the percentage is favorable. The advantages of the proposed method are demonstrated through comparing the results from two conventional data-driven fault diagnosis methods, i.e., SVM-based fault diagnosis method and the back-propagation (BP) network-based fault diagnosis method. The parameter σ in SVM was optimized by the grid search method. The three-layer BP network was used, and it thresholds and weights were determined by the genetic algorithm to seek the global optimal solution. The results of the comparative study are presented in **Table 2**. For the data set A, the training and testing accuracy of the BP network-based fault diagnosis method are higher than those of the SVM-based fault diagnosis method. As opposed to the data set A, the SVM-based fault diagnosis method has a high training and testing accuracy for the data set B, whereas the BP network-based fault diagnosis method is inferior for the data set B and its accuracy is less than 90%. For the data set C, both the SVM-based fault diagnosis method and the BP network-based fault diagnosis method exhibit a relatively high accuracy. As seen from **Table 2**, the proposed method is superior to the two conventional methods on all the three data sets, and its accuracy reaches 100%.

Data set	**SVM (%)**		**BP network (%)**		**The proposed method (%)**	
	Training	**Testing**	**Training**	**Testing**	**Training**	**Testing**
A	95.25	92.14	**100**	99.62	**100**	**100**
B	**100**	98.10	89.29	87.14	**100**	**100**
C	**100**	97.86	**100**	96.19	**100**	**100**

Table 2. The classification accuracy of the three methods.

4. The proposed heath assessment approach

Following the similar framework as Section 3, the procedures of the proposed data-driven health assessment approach for rotating machinery are presented in **Figure 8**. Instead of using the distance-based feature selection method, which is a supervised feature selection approach and needs to set the number of the states to discriminate, the principle component analysis (PCA) as an unsupervised feature selection tool is used here. The KDE and KLID are used together to construct a new health indicator, reflecting the health condition of the monitored rotating machinery.

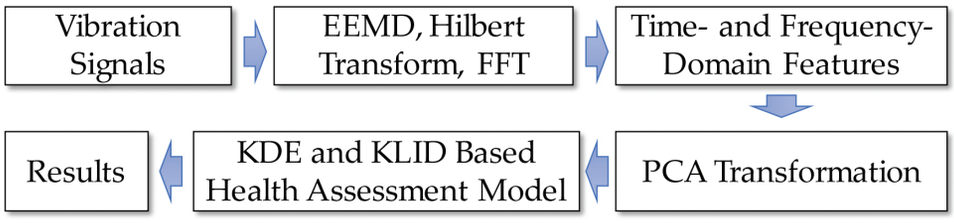


Figure 8. The procedures of the proposed health assessment approach for rotating machinery.

4.1. The principle component analysis (PCA)

The PCA proposed by Pearson [22] is a statistical procedure aiming to extract the directions with strong variability in a data set, and it can convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables. As its good capability in terms of reducing the dimensionality of data set, the PCA has been extensively used to deal with multivariate data in the field of pattern recognition, image processing, etc.

Mathematically, given a set of p dimensional feature vectors $\mathbf{x}_i (i = 1, 2, \ldots, n)$, the corresponding covariance matrix of the feature vectors can be computed by:

$$\mathbf{C} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^{\mathrm{T}}, \tag{13}$$

where

$$\boldsymbol{\mu} = \frac{1}{n}\sum_{t=1}^{n}\mathbf{x}_i \tag{14}$$

The principal components (PCs) can be computed by solving the eigenvalue and eigenvector of the covariance matrix $\mathbf{C}$ as follows:

$$\mathbf{C}\boldsymbol{\nu} = \boldsymbol{\lambda}\boldsymbol{\nu} \tag{15}$$

where $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_p]$ are the eigenvalues of the covariance matrix $\mathbf{C}$ in a descending order and $\mathbf{v} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p]$ are the associated eigenvectors.

To represent the original feature vector through a lower dimensional feature vector, the first m ($m \leq p$) eigenvectors that correspond to the first k largest eigenvalues will be selected. Oftentimes, a pre-specified threshold $\theta (\theta \in [0,1])$ needs to be given for a particular problem by user to satisfy:

$$\sum_{i=1}^{m}\lambda_i \Big/ \sum_{i=1}^{p}\lambda_i \geq \theta \tag{16}$$

A greater value of θ indicates to maintain a higher accuracy of the original feature vectors, and thus, more eigenvectors will be included. By this way, the number of eigenvectors for a particular problem can be determined so as to maintain the desired accuracy. The m dimensional feature vectors can be formulated as:

$$\boldsymbol{\varepsilon}_j = \mathbf{V}^{\mathrm{T}}\mathbf{x}_j, j = 1, 2, \ldots, n \tag{17}$$

By using the PCA, the dimensionality of the original feature vectors can be significantly reduced. The importance of the new feature vectors is denoted as $\mathbf{F} = (F_1, F_2, \ldots, F_m)$, and the importance of each new feature i can be evaluated by:

$$F_i = \lambda_i / \sum_{j=1}^{m} \lambda_j, \quad (i = 1, 2, \cdots, m) \tag{18}$$

4.2. The procedure of the proposed health assessment approach

The key idea behind the proposed health assessment approach is that the statistical characteristics of the samples at the good condition would exhibit an apparent discrepancy with that of the samples at the abnormal condition. In our work, the statistical characteristics of the samples are characterized by the KDE, whereas the KLID provides a quantitative way to measure the statistical discrepancy of the online monitoring samples with respect to the reference samples that are collected when the monitored device is at its good condition.

The overall flowchart of the proposed health assessment approach is shown in **Figure 9**. As shown in **Figure 9**, the features sensitive to the health status of rotating machinery are chosen. By conducting the PCA, the dimensionality of the selected features can be further reduced so as to reduce the computational burden in the ensuing steps. A moving window with width k will be used to dynamically construct a set of samples to evaluate the health condition of the monitored rotating machinery. An illustration of constructing sample sets over time through the moving window is delineated in **Figure 10**. With the assumption that the rotating machinery is at its good condition at the early stage of use, the samples collected by the moving window at the beginning of use will serve as the reference samples, whereas the samples collected by the moving window at the later stage will be statistically compared with the reference samples.

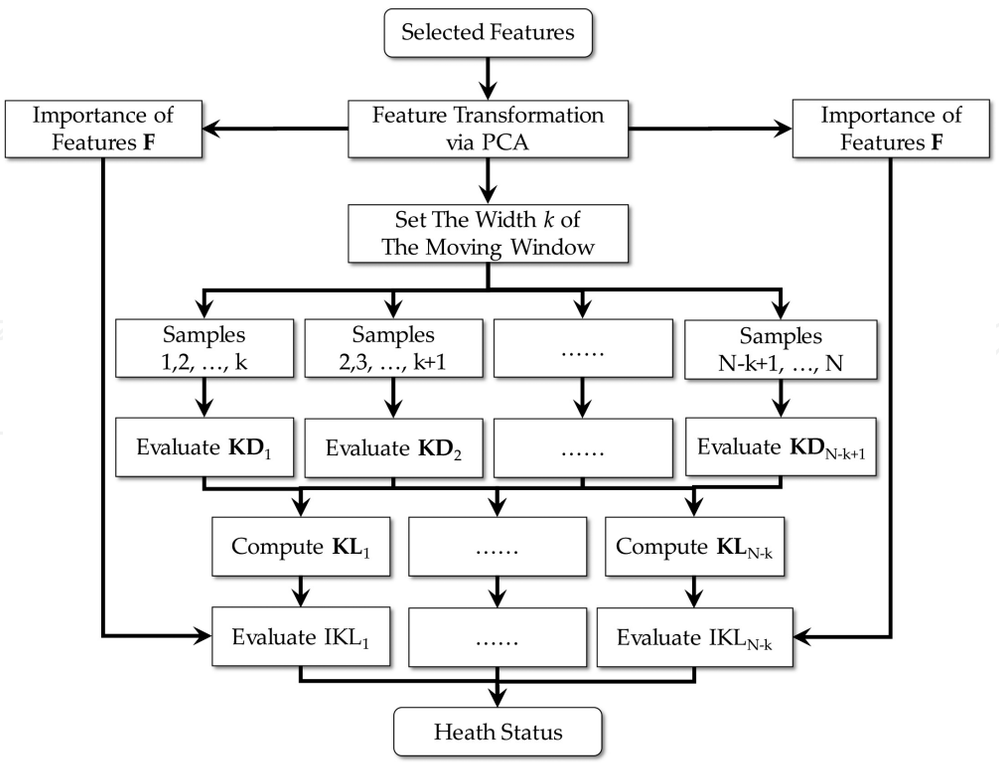


Figure 9. The flowchart of the proposed health assessment approach.

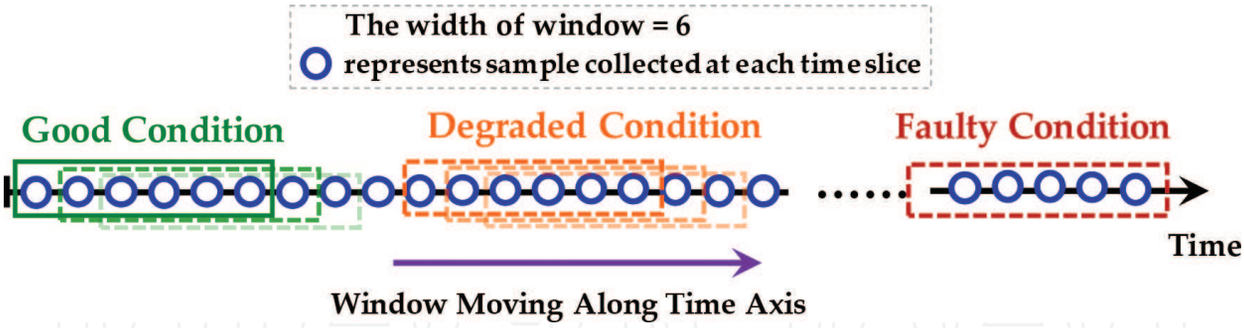


Figure 10. An illustration of the *k*-width moving window.

The statistical characteristic of each sample set in the moving window is characterized by the kernel density function, and it is denoted as KD_i^j for the jth ($j = 1, 2, \ldots, m$) PCA feature at the ith ($i = 1, 2, \ldots, N - k + 1$) window. $\mathbf{KD}_i = \left(KD_i^1, KD_i^2, \cdots, KD_i^m\right)$ is a collection of all the PCA features at the ith ($i = 1, 2, \ldots, N - k + 1$) window. The statistical discrepancy between the samples at the ith ($i = 1, 2, \ldots, N - k + 1$) window and $\mathbf{KD}_1 = \left(KD_1^1, KD_1^2, \cdots, KD_1^m\right)$ is quantified by the KLID and denoted as $\mathbf{KL}_i = \left(KL_i^1, KL_i^2, \cdots, KL_i^m\right)$ where KL_i^j computed by Eq. (10) is the KLID of the jth ($j = 1, 2, \ldots, m$) PCA feature at the ith ($i = 1, 2, \ldots, N - k + 1$) window. By taking into account the importance of the PCA features, the integrated KLID, denoted as IKL_i ($i = 1, 2, \ldots, N - k$), can be evaluated by Eq. (12), where $F_i (i = 1, 2, \ldots, m)$ in Eq. (12) takes values from Eq. (18). It should be noted that in the proposed approach IKL_i act as the health indicator for rotating machinery. A smaller value of IKL_i represents that the condition of the monitored device is close to the normal condition. On the contrary, if the condition of the monitored device gradually deviates the normal condition due to defects or faults, IKL_i will become a greater value.

4.3. A case study

To validate the effectiveness of the proposed method in terms of assessing the health status of rotating machinery, a case study for rolling element bearing is presented in this section.

4.3.1. Experimental setup

The experimental data are from the intelligent maintenance system (IMS) at the University of Cincinnati [23]. The run-to-failure data were collected from the experimental rig as shown in **Figure 11**, where the rolling bearings were working under a constant load condition. The rolling bearing test rig hosts four test Rexnord ZA-2115 double row rolling bearings on one shaft. Each row of the rolling bearings has 16 rollers, the section diameter is 71.5 mm, the rolling diameter is 8.4 mm, and the contacting angle of the roller is 15.17°. The rotation speed was set to be 2000 rpm. The sampling rate was 20 kHz, whereas the data length was 20,480 points. Three testing (i.e., Testings 1, 2, and 3) with identical rolling bearings were executed on this experimental rig.

In the reported experiment, three run-to-failure data sets were collected. At the end of Testing 1, an inner race defect occurred on Bearing 3. Bearing 4 developed a roller defect. At the end of

http://dx.doi.org/10.5772/67360

Testing 2, an outer race defect was found on Bearing 1. At the end of Testing 3, an outer race defect happened on Bearing 3. In our study, the run-to-failure data sets from Testings 1 and 2 are used to validate our proposed approach.

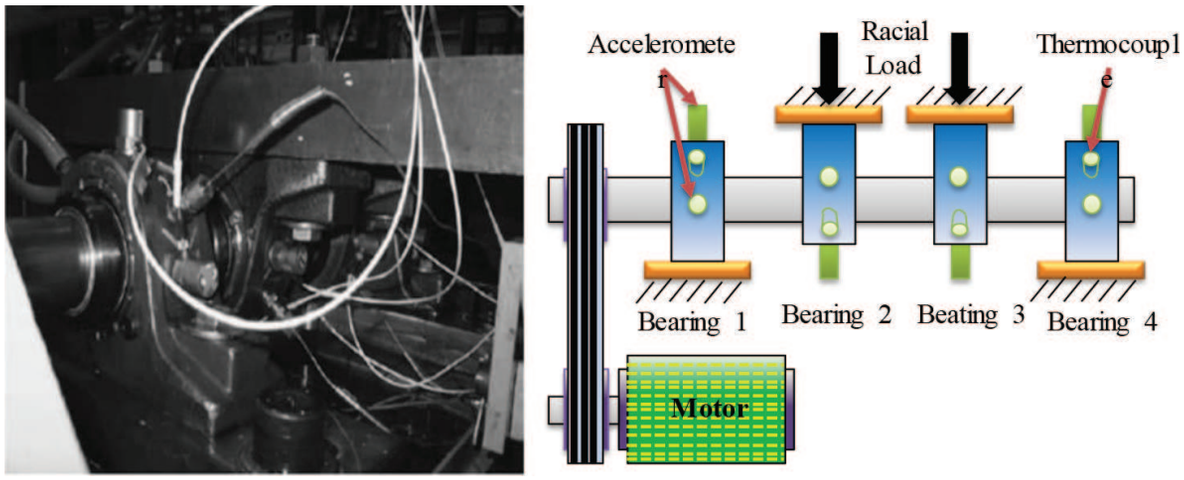


Figure 11. The rolling bearing experimental rig [23].

4.3.2. Results and analysis

In our study, the tools, like the ensemble empirical mode decomposition (EEMD), were first used to extract 95 representative features from the raw data sets. These features were transformed by the PCA to reduce the dimensionality. The importance of the principle components for Bearing 3 in Testing 1 and Bearing 1 in Testing 2 is shown in **Figure 12(a)** and **(b)**.

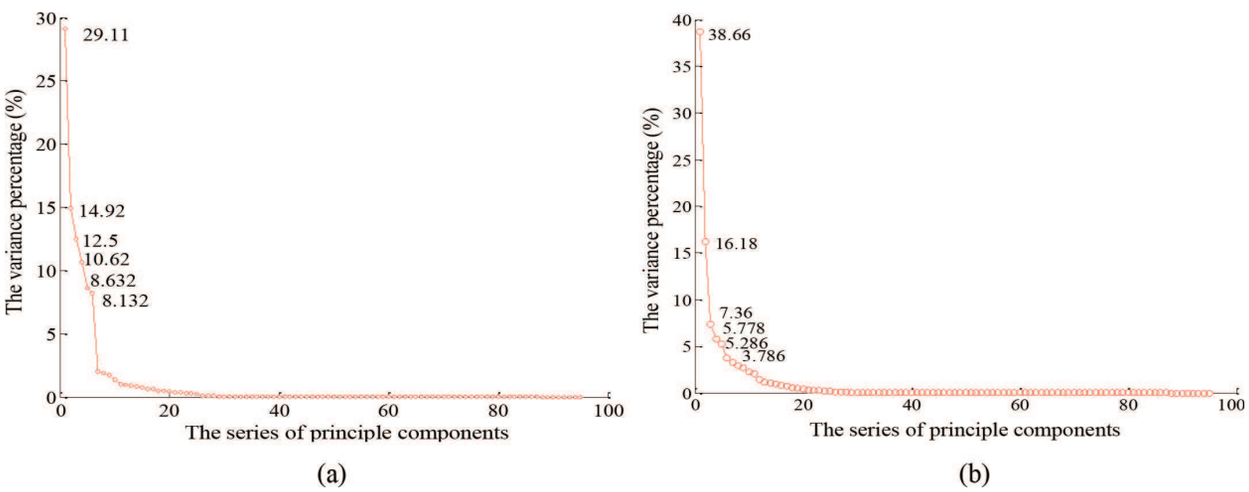


Figure 12. The importance of the principle components. (a) Bearing 3 in Testing 1 and (b) Bearing 1 in Testing 2.

From **Figure 12**, one can see that by using the first 10 principal components only, 90% accuracy can be maintained. On the other hand, with such a small sacrifice of accuracy, the dimensionality of features can be dramatically reduced from 95 to 10. Therefore, the first 10 principal

components were used as selected features and put into the proposed health assessment approach. The integrated KLID values from the proposed approach are plotted in **Figure 12 (a)** and **(b)** for Bearing 3 in Testing 1 and Bearing 1 in Testing 2, respectively, and these curves acting as the health indicator reflect the health status of the monitored bearings.

As shown in **Figure 13(a)**, the health indicator of Bearing 3 in Testing 1 has slight fluctuations at the early stage of the experiment. At the 1750th point, the health indicator rose up steeply and reached a great value in a short period of time (at the 1800th point). Such observation could indicate that the bearing put into use had a small manufacturing defects or a slight damage. It, therefore, leaded to the slight fluctuations of the health status at the beginning of the experiment. The tiny defect became serious suddenly at the 1750th points.

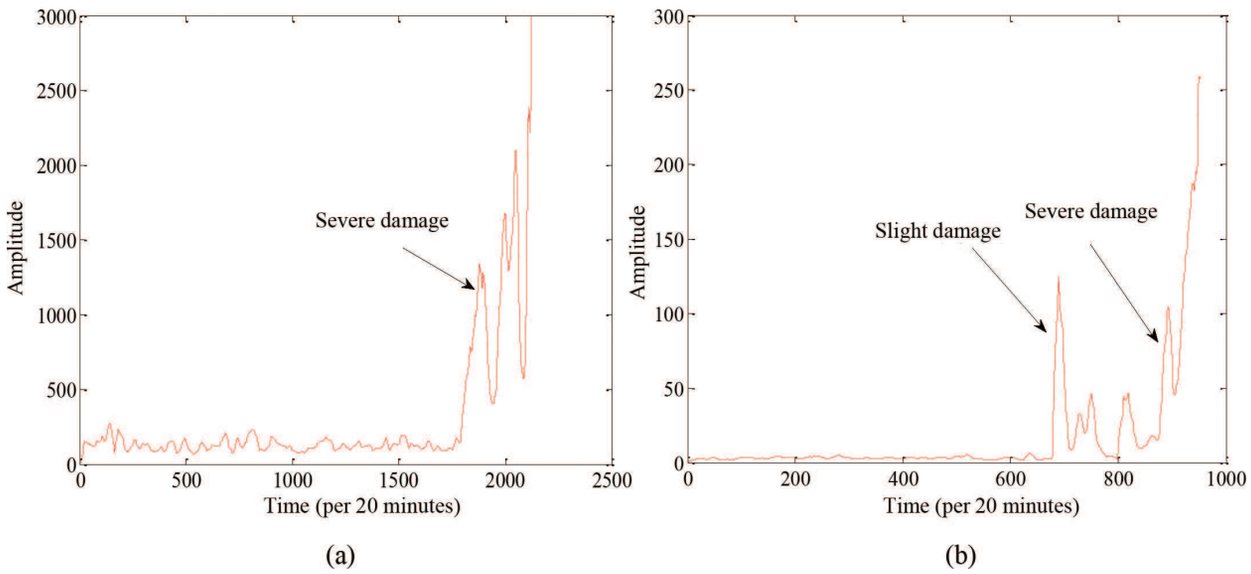


Figure 13. The results of health assessment. (a) Bearing 3 in Testing 1 and (b) Bearing 1 in Testing 2.

In **Figure 13(b)**, the health indicator of Bearing 1 in Testing 2 experienced two stages, namely the slight damage stage and the severe damage stage. Due to the slight damage, the health curve went up at the 680th point and reached the first peak value around the 700th point. However, as the bearing entered a new unhealthy but stable state, the health status of the bearing has a slight improvement as one can observed that the health curve dropped down for a while. Two hundred points (around 70 hours) later, the bearing jumped into the severe damage stage as shown in **Figure 13(b)**.

To illustrate the effectiveness of the proposed approach, the results from a recent literature are compared. The Locality Preserving Projection and Gaussian mixture model (GMM) was developed in Ref. [24] to construct a health assessment model for Bearing 1 in Testing 2. As found in Ref. [24], the health curve changed after the 700th point, indicating the occurrence of a slight damage. However, our proposed health indicator which rose up at the 680th point and reached the first maximum value at the 700th point. To examine the status of Bearing 1 in Testing 2 at the 680th point, the empirical mode decomposition (EMD) [25] was used to decompose the collected vibration data to four levels. The Hilbert transformation (HT) was

http://dx.doi.org/10.5772/67360

performed on the four intrinsic mode functions (IMFs). For more details about HT, please refer to Ref. [20]. The corresponding results are shown in **Figure 14**. From **Figure 14(b)**, one can easily find the frequency component of 236.3 Hz and its 2–4 times frequency components. On the other hand, the ball bass frequency at outer race (BPFO) can be theoretically calculated as follows:

$$f_{BPFO} = \frac{1}{2} \times n \times [1 - (d/D) \times \cos\varphi] \times N \quad = 236.4\text{Hz} \tag{19}$$

Thereby, one can conclude that the outer race defect occurred at the 680th point. Furthermore, this result illustrates that the proposed health assessment approach has a better capability to detect the incipient defect than the method proposed in Ref. [24].

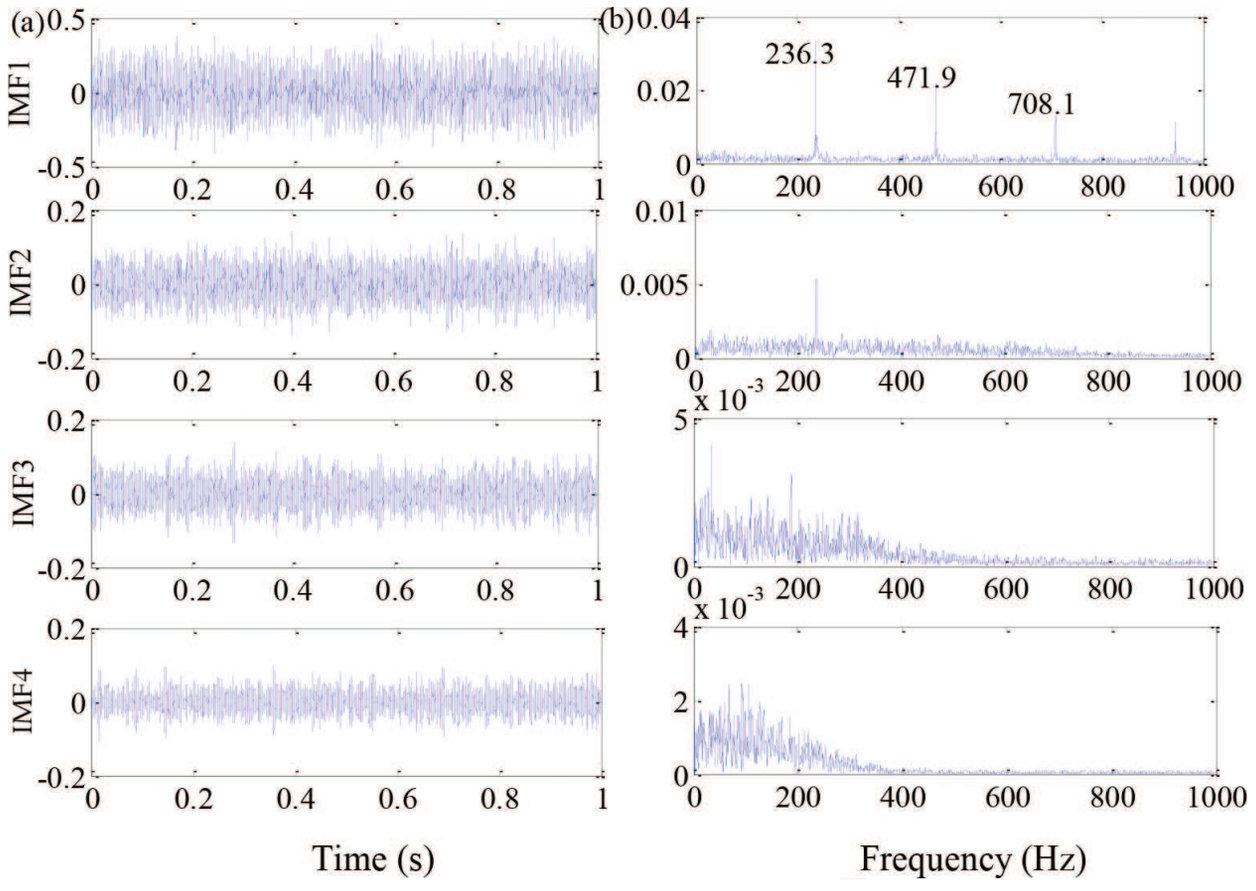


Figure 14. The results of the EMD and the corresponding Hilbert spectrum of Bearing 1 in Testing 2 at the 680th point. (a) The first four IMFs and (b) the Hilbert spectrum.

5. Conclusion

In this chapter, based on the kernel density estimation and the Kullback-Leibler divergence, a new data-driven fault diagnosis approach and a new health assessment approach are developed

by examining the statistical characteristics of the collected sample sets. By using the KDE and the KLID, the fault types or health status can be identified by comparing the integrated KLID of selected features. As demonstrated in the fault diagnosis examples, the proposed fault diagnosis approach has an exceptional performance on faulty pattern recognition, and it outperforms the conventional SVM-based and BP network-based methods. Meanwhile, in the example of health assessment, the proposed health assessment approach, which takes account of the statistical characteristics of sample sets, is capable of quantitatively tracking the health condition of the monitored rotating machinery.

Acknowledgement

The authors greatly acknowledge the grant support from the National Natural Science Foundation of China under contract number 71371042 and the Fundamental Research Funds for the Central Universities under contract number ZYGX2015J082.

Author details

Yu Liu*, Chen-Yao Yan and Fan Zhang

*Address all correspondence to: yuliu@uestc.edu.cn

School of Mechatronics Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, P.R. China

References

[1] Jardine A.K.S., Lin D., Banjevic D.. A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing. 2006;**20**(7):1483 –1510.

[2] Heng A., Zhang S., Tan A.C.C., Mathew J.. Rotating machinery prognostics: state of the art, challenges and opportunities. Mechanical Systems and Signal Processing. 2009;**23** (3):724–739.

[3] Chen J., Patton R.J.. Robust Model-Based Fault Diagnosis for Dynamic Systems. New York: Springer; 1999.

[4] Sekhar A.S.. Model-based identification of two cracks in a rotor system. Mechanical Systems and Signal Processing. 2004;**18**(4):977–983.

[5] Li Y., Kurfess T.R., Liang S.Y.. Stochastic prognostics for rolling element bearings. Mechanical Systems and Signal Processing. 2000;**14**(5):747–762.

[6] Tibaduiza D.A., Torres-Arredondo M.A., Mujica L.E., et al. A study of two unsupervised data driven statistical methodologies for detecting and classifying damages in structural health monitoring. Mechanical Systems and Signal Processing. 2013;**41**(1):467–484.

[7] Lei Y.G., He Z.J., Zi Y.Y.. A new approach to intelligent fault diagnosis of rotating machinery. Expert Systems with Applications. 2008;**35**(4):1593–1600.

[8] Cui H., Zhang L., Kang R., Lan X.. Research on fault diagnosis for reciprocating compressor valve using information entropy and SVM method. Journal of Loss Prevention in the Process Industries. 2009;**22**(6):864–867.

[9] Saravanan N., Ramachandran K.I.. Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN). Expert Systems with Applications. 2010;**37**(6):4168–4181.

[10] Widodo A., Yang B.S.. Support vector machine in machine condition monitoring and fault diagnosis. Mechanical Systems and Signal Processing. 2007;**21**(6):2560–2574.

[11] Rafiee J., Arvani F., Harifi A., Sadeghi M.H.. Intelligent condition monitoring of a gearbox using artificial neural network. Mechanical Systems and Signal Processing. 2007;**21** (4):1746–1754.

[12] Zhao F., Tian Z., Zeng Y.. Uncertainty quantification in gear remaining useful life prediction through an integrated prognostics method. IEEE Transactions on Reliability. 2013;**62** (1):146–159.

[13] Yang B.S., Kim K.J.. Application of Dempster–Shafer theory in fault diagnosis of induction motors using vibration and current signals. Mechanical Systems and Signal Processing. 2006;**20**(2):403–420.

[14] Rosenblatt M.. Remarks on some nonparametric estimates of a density function. The Annals of Mathematical Statistics. 1956;**27**(3):832–837.

[15] Parzen E.. On estimation of a probability density function and mode. The Annals of Mathematical Statistics. 1962;**33**(3):1065–1076.

[16] Kullback S., Leibler R.A.. On information and sufficiency. The Annals of Mathematical Statistics. 1951;**22**(1):79–86.

[17] Wu Z.H., Huang N.E.. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Advances in Adaptive Data Analysis. 2009;**1**(1):1–41.

[18] Lei Y.G., He Z.J., Zi Y.Y.. Application of the EEMD method to rotor fault diagnosis of rotating machinery. Mechanical Systems and Signal Processing. 2009;**23**(4):1327–1338.

[19] Yang B.S., Han T., An J.L.. ART–KOHONEN neural network for fault diagnosis of rotating machinery. Mechanical Systems and Signal Processing. 2004;**18**(3):645-657. Epanechnikov V.A. Non-parametric estimation of a multivariate probability density. Theory of Probability & Its Applications. 1969;**14**(1):153–158.

[20] Zhang F., Liu Y., Chen C., Li Y., Huang H.Z.. Fault diagnosis of rotating machinery based on kernel density estimation and Kullback-Leibler divergence. Journal of Mechanical Science and Technology. 2014;**28**(11):4441–4454.

[21] Bearing Data Center, Case Western Reserve University. Available from: http://www.eecs.cwru.edu/laboratory/Bearing [Accessed: June, 2009].

[22] Pearson K.. On lines and planes of closest fit to systems of points in space. Philosophical Magazine. 1901;**2**(11):559–572.

[23] Lee J., Qiu H., Yu J., Lin J.. Rexnord Technical Services. Bearing Data Set [Internet]. Available from: http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/ [Accessed: March 2014]

[24] Yu J.. Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models. Mechanical Systems and Signal Processing. 2011;**25**(7):2573–2588.

[25] Huang N.E., Shen Z., Long S.R., et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London 1971, **454**:903–995.

IntechOpen