*Research Article*

# Bearing Degradation Process Prediction Based on the Support Vector Machine and Markov Model

## Shaojiang Dong,[1] Shirong Yin,[1] Baoping Tang,[2] Lili Chen,[1] and Tianhong Luo[1,3]

[1] *School of Mechatronics and Automotive Engineering, Chongqing Jiaotong University, Chongqing 400074, China*
[2] *The State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing 400030, China*
[3] *Key Laboratory of Road Construction Technology and Equipment, Ministry of Education, Chang'an University, Xi'an 710021, China*

Correspondence should be addressed to Shaojiang Dong; dongshaojiang100@163.com

Predicting the degradation process of bearings before they reach the failure threshold is extremely important in industry. This paper proposed a novel method based on the support vector machine (SVM) and the Markov model to achieve this goal. Firstly, the features are extracted by time and time-frequency domain methods. However, the extracted original features are still with high dimensional and include superfluous information, and the nonlinear multifeatures fusion technique LTSA is used to merge the features and reduces the dimension. Then, based on the extracted features, the SVM model is used to predict the bearings degradation process, and the CAO method is used to determine the embedding dimension of the SVM model. After the bearing degradation process is predicted by SVM model, the Markov model is used to improve the prediction accuracy. The proposed method was validated by two bearing run-to-failure experiments, and the results proved the effectiveness of the methodology.

## 1. Introduction

Bearing is one of the most important components in rotating machinery. Accurate bearing degradation process prediction is the key to effective implement of condition based maintenance and can prevent unexpected failures and minimize overall maintenance costs [1, 2].

To achieve effective degradation process prediction of the bearing, firstly, the features should be extracted from the collected vibration data. Then, based on the extracted features effectively prediction models should be selected [3]. Feature extraction is the process of transforming the raw vibration data collected from running equipment to relevant information of health condition. There are three types of methods to deal with the raw vibration data: time domain analysis, frequency domain analysis, and time-frequency domain analysis. The three types of methods are often chosen to extract the feature. For example, Yu [4] chose the time domain and the frequency domain transform to describe the characteristics of the vibration signals. Yan et al. [5] chose the short-time Fourier transform to extract the features. Ocak et al. [6] chose the wavelet packet transform to extract the

feature of bearing wear information. Because the frequency features from FFT analysis results often tend to average out transient vibrations and thus not providing a wholesome measure of the bearing health status, in this paper, the time domain and the time-frequency domain characteristics are used to extract the original features.

Although the original features can be extracted, they are still with high dimension and include superfluous information. So the original features fusion and dimensional reduction method should be used to deal with the original features so as to select the typical features. The most commonly used features fusion and dimensional reduction method is principal component analysis (PCA) [7, 8]. But the PCA is mainly used for dealing with the linear data set, while the bearing vibration features are usually suppressed by the nonlinear characteristic features, so the PCA cannot work effectively. Therefore, it is a challenging to find an effective nonlinear features fusion and dimensional reduction method. In this research a new feature extraction method local tangent space alignment (LTSA) [9] is chosen. The LTSA is an efficient manifold-learning algorithm, which can be

used as a preprocessing method to transform the high dimensional data into more easily handled low dimensional data [10]; the method has been used in many fields, such as face recognition, character recognition, and image recognition [11, 12]. In this paper, the LTSA is used to achieve extracting the more sensitive features.

After selecting the typical features, another challenge is how to effectively predict the bearing degradation process based on the extracted features. The existing equipment degradation process prediction methods can be roughly classified into model-based (or physics-based modals) and data-driven methods [13]. The model-based methods predict the equipment degradation process using the physical models of the components and damage propagation models based on damage mechanics [14, 15]. However, equipment dynamic response and damage propagation processes are typically very complex, and authentic physics-based models are very difficult to build [16]. Data-driven methods, also known as artificial intelligent approaches, are derived directly from routine condition monitoring data of the monitored system, which predicts the failure progression based on the learning or training process. The more prior the data is used for the training process, the more accurate is the model obtained [17]. Artificial intelligent techniques have been increasingly applied to bearing remaining life prediction recently. Lee et al. [18] presented an Elman neural network method for health condition prediction. Huang et al. [19] proposed a back-propagation network-based method for bearing degradation process prediction. However, the neural networks have the drawbacks of slow convergence; difficulty in escaping from local minima; uncertain network structure, especially when doing the bearing degradation process prediction problem with large data, and those problems will be more troublesome. The SVM [20] is most widely used recently and has a good identify and regression ability. In this paper, the SVM is used to predict the bearings degradation process.

Although the SVM is effective in predicting the bearing running state, the prediction error still exists. Because any prediction methods based on the historical data for future prediction will more or less have some prediction error, it is necessary to improve the prediction results. However, the prediction error has the character of being affected by many factors, fluctuations, showing a great random, and the error points are not related. So if we want to achieve high prediction accuracy, we need to find the discipline of the prediction error and correct the error of prediction. The Markov model [21] used the state transition matrix to achieve a more precise prediction. It can be used to achieve the improvement of the prediction accuracy. So in this research, the Markov model is used to improve the prediction accuracy.

The remainder of this paper is organized as follows. The methods of features extraction and the theory of dimensional reduction method LTSA are introduced in Section 2. The SVM model and the Markov model for bearing degradation process prediction are described in Section 3. In Section 4, the flowchart and the procedure of this research are introduced. The case validation and actual application are presented in Section 5. Finally, the conclusions are given in Section 6.

## 2. Methods of Signal Processing and Dimensional Reduction

*2.1. Feature Extraction.* This section presents a brief discussion on original feature extraction from time domain, time-frequency domain of vibration signals. Time domain methods usually involve statistical features that are sensitive to impulsive oscillation, such as kurtosis, skewness, peak-peak (P-P), RMS, and sample variance. The 5 domain statistical features are used as original features in time domain has been used in the literatures [2, 3]:

$$\text{RMS} = \sqrt{\frac{\sum x_i^2}{N}}, \tag{1}$$

where $N$ is the number of discrete points and $x_i$ represents the signal value at those points,

$$\text{Variance} = s^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} (x_i - \overline{x})^2, \tag{2}$$

where $\overline{x}$ is the mean value.

The $i$th central moment for a set of data is defined as

$$M_i = \frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^i. \tag{3}$$

The normalized forth moment, kurtosis, which is commonly used in bearing diagnostics, is defined as

$$\text{Kurtosis} = \frac{M_4}{(M_2)^2} - 3. \tag{4}$$

The skewness is defined as

$$\text{Skewness} = \frac{M_3}{(M_2)^{3/2}}. \tag{5}$$

The peak-peak is defined as

$$\text{peak-peak} = \max(x_i) - \min(x_i). \tag{6}$$

Empirical mode decomposition (EMD) is a powerful tool in time-frequency domain analysis. The advantage of EMD is the presentation of signals in time-frequency distribution diagrams with multiresolution, during which choosing some parameters is not needed. This property is essential in the detection of bearing faults. The EMD energy can represent the characteristic of vibration signals, and thus it is used as the input features. The (intrinsic mode function) IMF energy data sets are chosen as original features in this paper. The original features for bearing degradation process prediction based on the original features are shown in Table 1.

*2.2. Dimensional Reduction Based on the LTSA.* Because the generated original feature sets are still with high dimension and include superfluous information, the feature extraction method LTSA is used to fuse the relevant useful features and extracts more sensitive features to work as the input of the proposed prediction model.

Table 1: Original features for bearing fault diagnosis and performance assessment.

| Domain | Feature |
| --- | --- |
| Time domain | RMS, Kurtosis, Skewness, Peak-Peak, and sample variance |
| Time-frequency domain | EMD energy $(F_{e1}, F_{e2}, F_{e3}, \ldots, F_{en})$ |

The basic idea of LTSA is to use the tangent space of sample points to represent the geometry of the local character. Then these local manifold structures of space are lined up to construct the global coordinates. Given a data set $X = [x_1, x_2, \ldots, x_N], x_i \in R^m$, a mainstream shape of $d$-dimension $(m > d)$ is extracted. The LTSA feature extraction algorithm is as follows [9].

(1) Extract local information: for each $x_i, i = 1, 2, \ldots, N$, use the Euclidean distance to determine a set $x_i = [x_{i,1}, x_{i,2}, \ldots, x_{i,k_i}]$ of its neighborhood adjacent points (e.g., $k$ the nearest neighbors).

(2) Local linear fitting: in the neighborhood of data points $x_i$, a set of orthogonal basis $Q_i$ can be selected to construct the $d$-dimension neighborhood space of $x_i$ and the orthogonal projection of each point $x_{i,j}(j = 1, 2, \ldots, N)$ can be calculated to the tangent space of $\theta_j^{(i)} = Q_i^T(x_{i,j} - \overline{x_i})$. $\overline{x_i}$ is the mean data for the neighborhood. The orthogonal projection in the tangent space of neighborhood data of $x_i$ is composed of local coordinate $\Theta_i = [\theta_{(i),1}, \theta_{(i),2}, \ldots, \theta_{(i),k_i}]$ that describes the most important information of the geometry of the $x_i$.

(3) Global order of the local coordinates: supposing that the global coordinates of $x_i$ converted by the $\Theta_i$ is $T_i = [t_{i1}, t_{i2}, \ldots, t_{ik_i}]$, and then the error is

$$E_i = T_i \left[ I - \left( \frac{1}{k} \right) ee^T \right] - L_i \Theta_i, \tag{7}$$

where the $I$ is the identity matrix; the $e$ is the unit vector; the $k$ is the points number of the neighborhood; the $L_i$ is the transformation matrix. In order to minimize the error, the $T_i$ and $L_i$ should be found, and then

$$L_i = T_i \left( I - \left( \frac{1}{k} \right) ee^T \right) \Theta_i^*,$$
$$E_i = T_i \left( I - \left( \frac{1}{k} \right) ee^T \right) (I - \Theta_i^* \Theta_i), \tag{8}$$

where the $\Theta_i^*$ is the Moor-Penrose generalized inverse of $\Theta_i$. Suppose

$$B = PWW^T P^T. \tag{9}$$

Let $P = [P_1, P_2, \ldots, P_N]$, $TP_i = T_i$, $P_i$ be a selected matrix from 0-1; the $T$ are global coordinates, and their weight matrix is

$$W = \text{diag}(W_1, W_2, \ldots, W_N),$$
$$W_i = \left( I - \left( \frac{1}{k} \right) ee^T \right) (I - \Theta_i^* \Theta_i). \tag{10}$$

The constraints is $TT^T = I_d$.

(4) Extract of the low-dimensional manifolds feature: since the $e$ is the eigenvalue of matrix $B$, the corresponding minimum eigenvectors matrix is composed of eigenvalue. The value of section 2 to section $d + 1$ of matrix $B$ make of the $T$. $T$ is the global coordinate mapping in the mainstream form of low-dimensional transformed from the nonlinear high-dimensional data set of $X$.

The procedure of feature extraction can be described as follow.

(1) Use the time domain analysis methods kurtosis, skewness, peak-peak, RMS, and sample variance to extract the statistical features.

(2) Use the EMD method to decompose the collected vibration signal of each data set and get the IMF components; calculate the energy of each IMF component and get the features of the bearing in this time; then get the features of the other data sets.

(3) Use the LTSA to reduce the original features dimensions and get the main features; the extracted features are used as input of the SVM model for bearing degradation process prediction.

## 3. The SVM and Markov Model for Degradation Process Prediction

*3.1. SVM Prediction Model.* SVM is a machine learning tool that uses statistical learning theory to solve multidimensional functions. It is based on structural risk minimization principles, which overcomes the extralearning problem of ANN.

The learning process of a SVM regression model is essentially a problem in quadratic programming. Given a set of data points $(x_1, y_1) \cdots (x_k, y_k)$ such that $x_i \in R^n$ as input and $y_i \in R$ as target output, the regression problem is to find a function such as

$$y = f(x) = w\phi(x) + b, \tag{11}$$

where $\phi(x)$ is the high dimensional feature space, which is nonlinear mapped from the input space $x$, $w$ is the weight vector, and $b$ is the bias [22].

After training, the corresponding $y$ can be found through $f(x)$ for the $x$ outside the sample. The $\varepsilon$-support vector regression ($\varepsilon$-SVR) by Vapnik controls the precision of the algorithm through a specified tolerance error $\varepsilon$. The error of the sample is $\xi$, regardless of the loss, when $|\xi| \leq \varepsilon$; else consider the loss as $|\xi| - \varepsilon$. First, map the sample into a high

dimensional feature space by a nonlinear mapping function and convert the problem of the nonlinear function estimates into a linear regression problem in a high dimensional feature space. If we let $\varphi(x)$ be the conversion function from the sample space into the high dimension feature space, then the problem of solving the parameters of $f(x)$ is converted to solving an optimization problem (12) with the constraints in (13):

$$\min \qquad \frac{1}{2}\|w\|^2 = \frac{1}{2}(w \cdot w) \qquad (12)$$

$$\text{Subject to} \quad y_i - (w \cdot \varphi(x) + b) \leq \varepsilon,$$

$$(w \cdot \varphi(x) + b) - y_i \leq \varepsilon, \qquad (13)$$

$$i = 1, 2, \ldots, l.$$

The feature space is one of high dimensionality and the target function is nondifferentiable. In general, the SVM regression problem is solved by establishing a Lagrange function and converting this problem to a dual optimization, that is, problem (14) with constraint of (15) in order to determine the Lagrange multipliers $\alpha_i, \alpha_i^*$,

$$\max \quad -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)(x_i, x_J)$$

$$\qquad (14)$$

$$-\varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l}y_i(\alpha_i - \alpha_i^*)$$

$$\text{Subject to} \quad \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0$$

$$\qquad (15)$$

$$\alpha_i, \alpha_i^* \in [0, C],$$

where $\alpha_i, \alpha_i^*$ are Lagrange multipliers and $\alpha_i, \alpha_i^* \geq 0$. $\alpha_i \times \alpha_i^* = 0$. $C$ evaluates the tradeoff between the empirical risk and the smoothness of the model.

The SVM regression problem has therefore been transformed into a quadratic programming problem. The regression equation can be obtained by solving this problem. With the kernel function $K(x_i, x_j)$, the corresponding regression function is provided by

$$f(x) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j) + b, \qquad (16)$$

where the kernel function $K(x_i, x_j)$ is an internal product of vectors $x_i$ and $x_j$ in feature spaces $\varphi(x_i)$ and $\varphi(x_j)$.

*3.2. The Prediction Strategy and the Structure of the SVM Model.* Traditional forecasting methods mainly achieve single-step prediction; when those methods are used for multisteps prediction, they cannot get an overall development trend of the series. Multisteps prediction method has the ability to obtain overall information of the series which provides the possibility for long-term prediction. There are two typical alternatives to build multisteps life prediction model. One is iterated prediction and the other is direct

prediction. The comparison of the two strategies can be found in a number of literatures [23]. Marcellino et al. [24] presented a large-scale empirical comparison of iterated versus direct prediction. The results show that iterated prediction typically outperforms the direct prediction. So, the iterated multisteps prediction strategy has numerous advantages and will be adopted in this paper.

In order to determine the structure of the SVM, we constructed a three layers SVM prediction model. But to achieve the multisteps time series life prediction a basic problem should be suppressed. That is how many essential observations (inputs) are used for forecasting the future value (the output node number is 1), so-called embedding dimension $d$. In order to suppress the problem, the CAO method [25], which is particularly efficient to determine the minimum embedding dimension through the expansion of neighbor point in the embedding space, is employed to select an appropriate embedding dimension $d$. Then, the SVM input node number is determined.

To effectively select an appropriate embedding dimension based on the CAO method, the phase space reconstruction method should be mentioned. The fundamental theorem of phase space reconstruction is pioneered by Takens [26]. For an $N$-point time series $\mathbf{X} = \{x_1, x_2, \ldots, x_N\}$, a sequence of vectors $y_i$ in a new space can be generated as $y_{i(d)} = \{x_i, x_{i+\tau}, \ldots, x_{i+(d-1)\tau}\}$, where $i = 1, 2, \ldots, N_d$, $N_d = N - (d-1)\tau$ is the length of the reconstructed vector $y_i$, $d$ is the embedding dimension of the reconstructed state space, and $\tau$ is embedding delay time. The time delay $\tau$ is chosen through the autocorrelation function [27]:

$$C(\tau) = \frac{\sum_{i=1}^{N-\tau} x_i' x_{i+\tau}'}{\sum_{i=1}^{N-\tau} (x_i')^2}, \qquad (17)$$

where $x_i' = x_i - \overline{x}$, $\overline{x}$ is the average value of the time series. The optimal time delay $\tau$ is determined when the first minimum value of $C(\tau)$ occurs.

The embedding dimension $d$ is chosen through CAO method, defining the quantity as follows:

$$a(i, d) = \frac{\|y_i(d+1) - y_{n(i,d)}(d+1)\|}{\|y_i(d) - y_{n(i,d)}(d)\|}, \qquad (18)$$

where $\|\cdot\|$ is the Euclidian distance and is given by the maximum norm. $y_i(d)$ means the $i$th reconstructed vector and $n(i, d)$ is an integer, so that $y_{n(i,d)}(d)$ is the nearest neighbor of $y_i(d)$ in the embedding dimension $d$. A new quantity is defined as the mean value of all $a(i, d)'s$:

$$E(d) = \frac{1}{N - d\tau}\sum_{i=1}^{N-d\tau} a(i, d), \qquad (19)$$

where $E(d)$ is only dependent on the dimension $d$ and time delay $\tau$. To investigate its variation from $d$ to $d+1$, the parameter $E_1$ is given by

$$E_1(d) = \frac{E(d+1)}{E(d)}. \qquad (20)$$

By increasing the value of $d$, the value $E_1(d)$ is also increased and it stops increasing when the time series comes from a deterministic process. If a plateau is observed for $d \geq d_0$, then $d_0 + 1$ is the minimum embedding dimension. But $E_1(d)$ has the problem of slowly increasing or has stopped changing if $d$ is sufficiently large. CAO introduced another quantity $E_2(d)$ to overcome the problem:

$$E_2(d) = \frac{E^*(d+1)}{E^*(d)}, \tag{21}$$

where

$$E^*(d) = \frac{1}{N - d\tau} \sum_{i=1}^{N-d\tau} |x_{i+d\tau} - x_{n(i,d)+d\tau}|. \tag{22}$$

Through CAO method, the embedding dimension $d$ of the SVM prediction model is chosen. The structure of the SVM model is determined.

### 3.3. SOM Clustering Method to Divide the Prediction Error.
State division is the process to determine the mapping from random variable to the state space. How to obtain state division is a crux for Markov model. Traditionally, it is performed by the state division approach described as follows. Let $X = \{x_1, x_2, \ldots, x_n\}$ be the random sequence; let $S = \{1, 2, \ldots, j, \ldots, m\}$ denote the state space; given $a_0 < a_1 < \cdots < a_m$ if random variable $x_i \in [a_{j-1}, a_j]$, where $1 \leq i \leq n$, $1 \leq j \leq m$, then the variable $x_i$ belongs to the state $j$, and the division of $[a_{j-1}, a_j]$ is usually uniform divided. However, the uniform divided method depends on the people's experience, which will affect the prediction precise. In this research, the SOM neural network [28] is used to divide the state. The SOM can be created from highly deviating, nonlinear data. After the data are input, the SOM is trained iteratively.

In each training step, one sample vector $X$ from the input data set is chosen randomly, and the distance between it and all the weight vectors of the SOM, which are originally initialised randomly, is calculated using some distance measure. The best matching unit (BMU) is the map unit, whose weight vector is closest to $X$. After the BMU is identified, the weight vectors of the BMU, as well as its topological neighbors, are updated so that they are moved closer to the input vector in the input space. The vectors are updated following the learning rule:

$$m_i(t+1) = m_i(t) + \alpha(t) h(n_{\text{BMU}}, n_i, t)(X - m_i(t)), \tag{23}$$

where $h(n_{\text{BMU}}, n_i, t)$ is the neighborhood function, which is monotonically decreasing with respect to the distance between the BMU $n_{\text{BMU}}$ and $n_i$ in the grid, and the training time $\alpha(t)$ is the learning rate; a decreasing function with $0 < \alpha(t) < 1$.

At the end of the learning process, the weight vectors are grouped into clusters depending on their distance in the input space. Unlike networks based on supervised learning, which require that target values corresponding to input vectors are known, the SOM can be used to cluster data without knowing the class membership of the input data. This character is suitable for the problem of the prediction error of the SVM

model which is not clear to us, so we should classify the error without prelearning, and therefore, the function of SOM method determined is an efficient and necessary method for clustering the state. Based on the clustering results, the state is divided into some districts.

### 3.4. Markov Prediction Model to Improve the Prediction Accuracy.
Consider a stochastic process $\{x_n, n = 1, 2, \ldots\}$ that takes on a finite or countable number of possible values. Unless otherwise mentioned, this set of possible values of the process will be denoted by the set of nonnegative integers $\{1, 2, \ldots\}$. If $x_n = i$, then the process is said to be in state $i$ at time $n$. Suppose that whenever the process is in state $i$; there is a fixed probability $p_{ij}$ that it will be next instating $j$. That is

$$p\{x_{n+1} = j \mid x_n = i, x_{n-1} = i_{n-1}, \ldots, x_1 = i_1\}$$
$$= p\{x_{n+1} = j \mid x_n = i\} = p_{ij} \tag{24}$$

for all states $i_0, i_1, \ldots, i_{n-1}, i, j$ and all $n \geq 0$. Such a stochastic process is known as a Markov chain. Equation (24) can be interpreted as stating that for a Markov model, the conditional distribution of any state $x_{n+1}$, given the past states $x_1, x_2, \ldots, x_{n-1}$ and the present state $x_n$, is independent of the past states and depends only on the present state. This is called the Markovian property. The value $p_{ij}$ represents the probability that the process will, when in state $i$, next make a transition into state $j$. Since probabilities are nonnegative and since the process must make a transition into some state, we have that

$$p_{ij} \geq 0, \quad i, j \geq 0; \qquad \sum_{j=1}^{\infty} p_{ij} = 1, \quad i = 1, 2, \ldots. \tag{25}$$

If the process has a finite number of states, which means the state space $S = \{1, 2, \ldots, i, j, \ldots, s\}$, then the Markov chain model can be defined by the matrix of one-step transition probabilities, denoted as

$$p = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & \vdots & \cdots & \vdots \\ p_{s1} & p_{s2} & \cdots & p_{ss} \end{bmatrix}. \tag{26}$$

The initial probability is computed by

$$p_{ij} = \frac{N_{ij}}{N_i}, \tag{27}$$

where $N_{ij}$ denotes the transition times from state $i$ to state $j$ and $N_i$ denotes the number of random variables $\{x_n, n = 1, 2, \ldots, m\}$ belonging to state $i$.

Markov model adopts state vector and state transition matrix to deal with the prediction issue. Suppose that the state vector of moment $t - 1$ is $p_{t-1}$, the state vector of moment $t$ is $p_t$, and the state transition matrix is $p$; then the relationship is

$$p_t = p_{t-1}p, \quad t = 1, 2, \ldots, K. \tag{28}$$

Update $t$ from 1 to $K$, and then
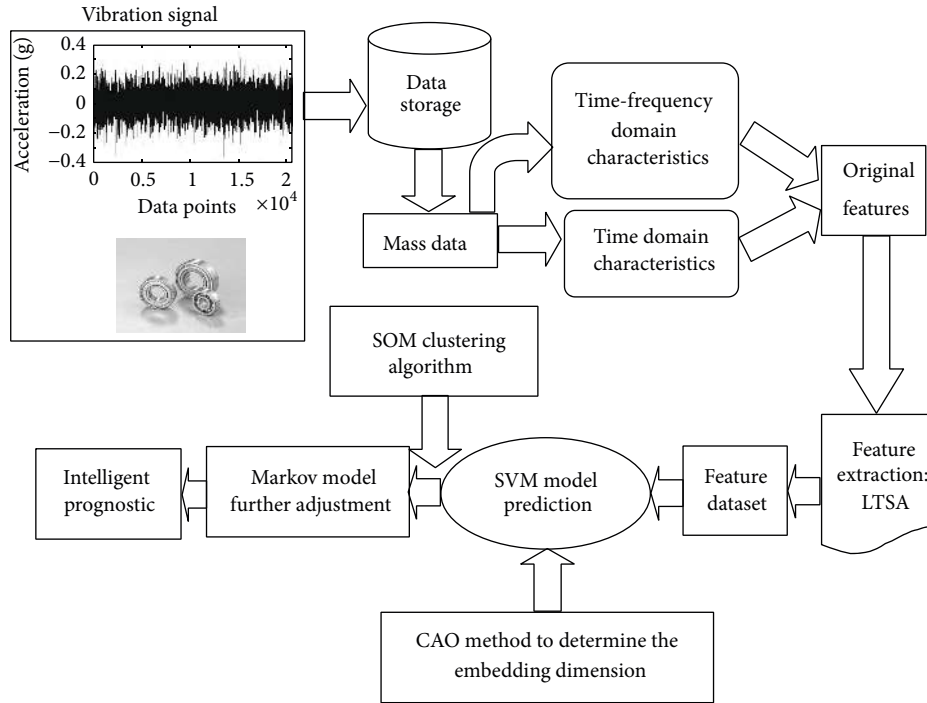
$$p_k = p_0 p^K, \tag{29}$$

FIGURE 1: The flowchart of the proposed method.

where $p^K$ is the state vector at moment $K$. Equation (29) is the basic Markov prediction model, if the initial state vector and the transition matrix are given, which allows calculation of any possible future state vector.

## 4. Proposed Method

The flowchart of the proposed method is shown in Figure 1. The method consists of four procedures sequentially: data processing and features extraction, merge of the original features, constructing-training SVM model and predicting, and Markov model for improving the prediction result. The role of each procedure is explained as follows.

*Step 1.* Data processing and features extraction. The time domain and time-frequency domain signal processing methods are used to extract the original features from the collected mass vibration data.

*Step 2.* Merge of the original features. The LTSA method is used to extract the typical features and reduce the dimension of the features. The extracted features are used for training the SVM model.

*Step 3.* Constructing the SVM model. The SVM model is constructed; the CAO method is used to determine the embedding dimension. The iterated multistep prediction method is used to forecast the future value.

*Step 4.* Markov model for improving the prediction result. This procedure uses the SOM method to cluster the prediction error before the Markov method; based on the state

division results the Markov model is used to improve the prediction results obtained by SVM model, to get a more precise prediction.

## 5. Validation and Application

*5.1. Validation.* In order to validate the effect of the proposed method, a validation test is proposed. The vibration signals used in this paper are provided by the Center for Intelligent Maintenance Systems (IMS), University of Cincinnati [29]. The experimental data sets are generated from bearing run-to-failure tests under constant load conditions on a specially designed test rig as shown in Figure 2. The rotation speed is kept constant at 2000 rpm. A radial load of 6000 lbs is added to the shaft and bearing by a spring mechanism. The data sampling rate is 20 kHz and the data length is 20,480 points as shown in Figure 3. It took a total of 7 days until the bearing fails. At the end, one bearing with serious wreck is used to test the proposed method as shown in Figure 4.

The time domain and time-frequency domain methods are used to deal with the collected vibration data as described in Section 2, Table 1. The measurements value of kurtosis and skewness are depicted in Figures 5(a) and 5(b).The IMF1 and IMF2 energy are depicted in Figures 6(a) and 6(b).

From the extracted features showed in Figures 5 and 6 we can see the following. (1) The bearing is in normal condition during the time correlated with the first 700 points. After that time, the condition of bearing suddenly changes. It indicates that there are some faults occurring in this bearing. (2) Different features reflect the bearing running state in different shapes. For example, the kurtosis and the IMF1
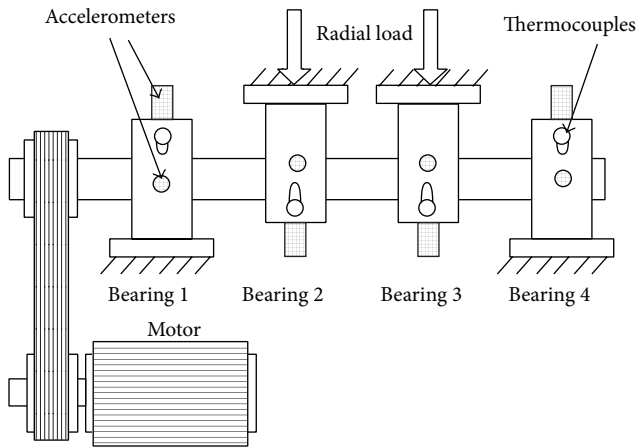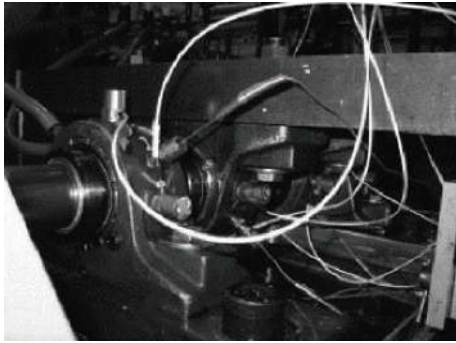
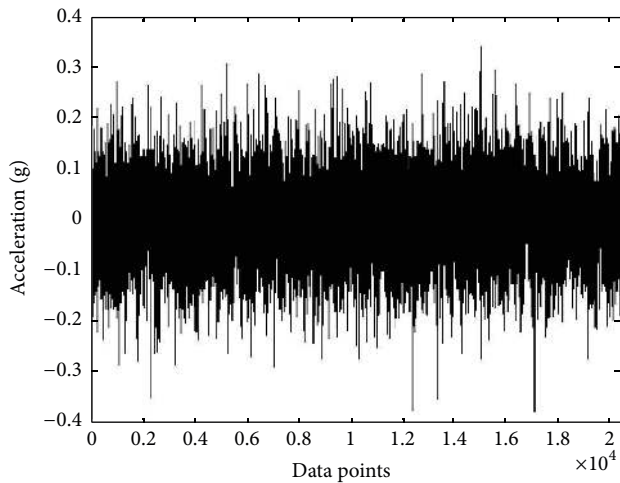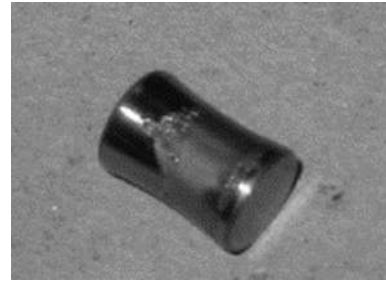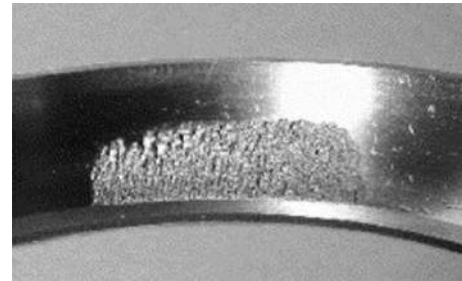FIGURE 2: Bearing test tig and sensor placement illustration [29].



FIGURE 3: Vibration signal.



(a)



(b)

FIGURE 4: The bearing components with serious wrecked after test with roller element defect and outer race defect.

indict the bearing running state with an upward trend, while the skewness indicts the bearing running state with a downward trend. The kurtosis, skewness, and the IMF1, indicting the bearing running state, have a large fluctuation after the 700-point, while the IMF2, indicting the bearing running state, have a large fluctuation till after the 850-point. So using these features is not appropriate to reflect the bearing running condition. So it is very necessary to extract a sensitive feature by the original features to appropriately reflect the bearing running condition.

We performed feature extraction by means of LTSA to extract a sensitive feature and reduce the dimensionality of calculated features. After LTSA is used (in this article the parameters of the neighborhood factor $k$ equals 8, the embedding dimension $d$ equals 1), the bearing running state features dataset is got. The first main projected vector is chosen as the input of the SVM model. The result is shown in Figure 7. In comparison with the LTSA method, we also extracted the features through the PCA method, and the first main principal component is chosen. The result is shown in Figure 8.

From Figures 7 and 8 we can see that the LTSA method can extract an effective feature dataset, which is sensitive to the changes of bearing running state, while the extracted feature is based on PCA method with a bad effect, before the 700 point; we even cannot see the fluctuation of the bearing running state, and the trend convert is also not obvious, from which we cannot know the bearing running state effectively. This result indicates that information extracted by LTSA could be more effective than that extracted by PCA.

After extracting the typical features, the CAO method is used to determine the embedding dimension of the SVM model, based on the theorem of phase space reconstruction. We first choose the delay time $\tau$ through the autocorrelation function.

The optimal time delay $\tau$ is determined when the first minimum value of $C(\tau)$ occurs.

Based on the extracted features dataset, the delay time $\tau$ is set to 3 for the projected vector values through the autocorrelation function as shown in Figure 9.

Then the embedding dimension is selected by the CAO method. The result is shown in Figure 10; the optimal embedding dimension $d$ for the projected vector is chosen as 10.
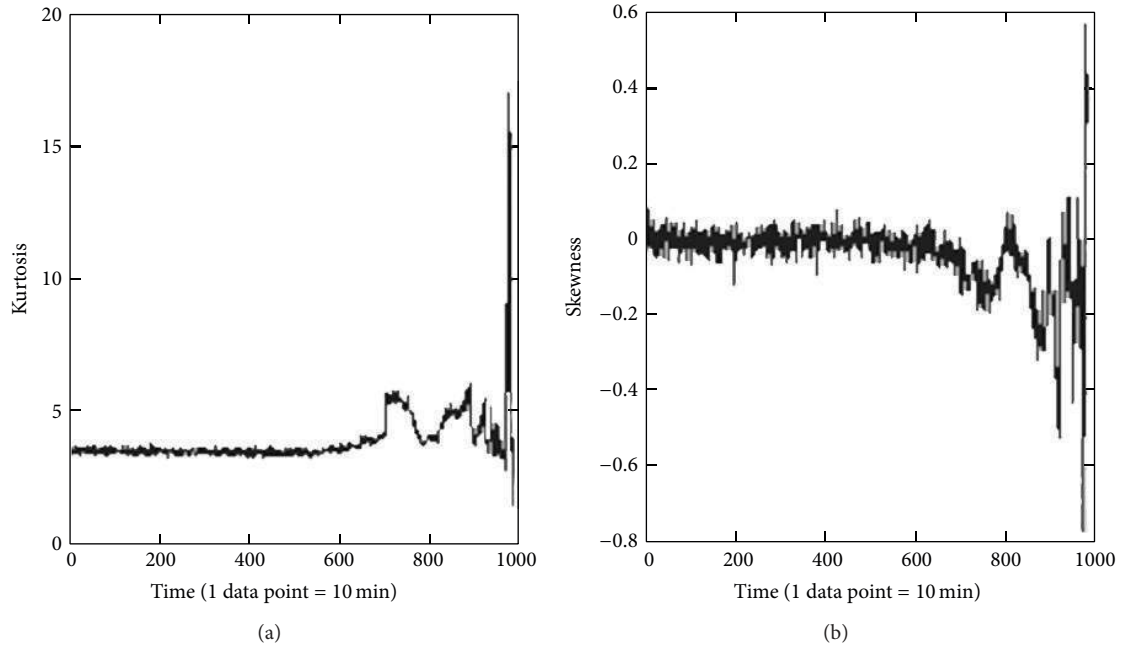
(a)



(b)

FIGURE 5: (a) Kurtosis value and (b) Skewness value of bearing vibration data.
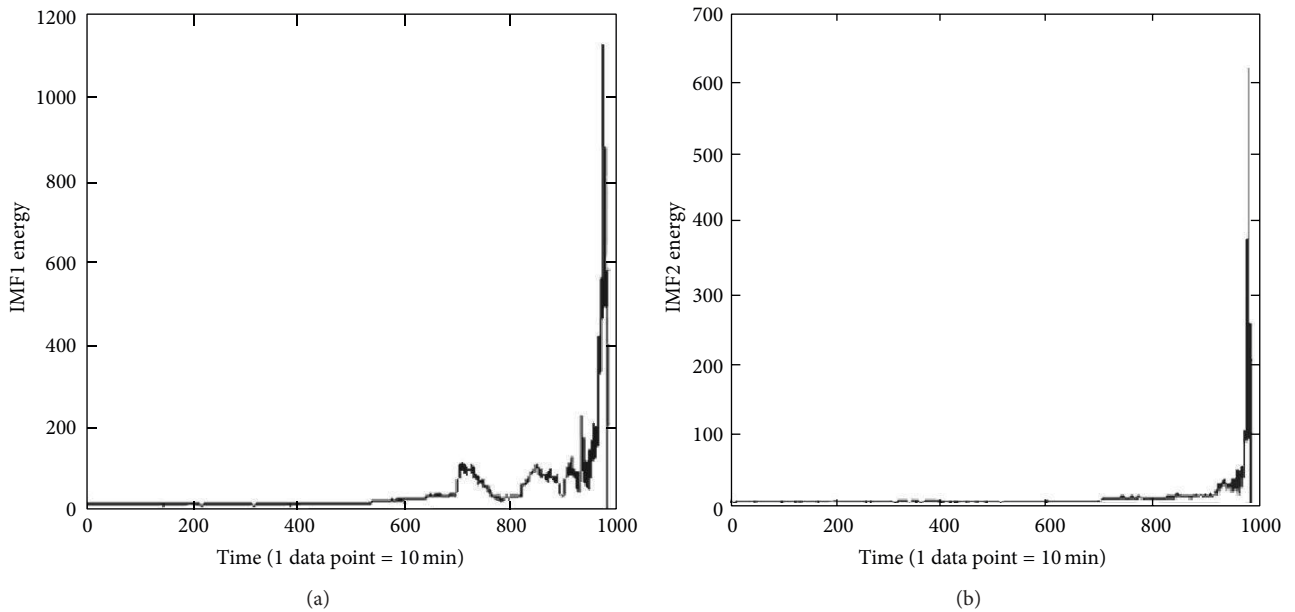


(a)



(b)

FIGURE 6: (a) IMF1 energy of bearing vibration data; (b) IMF2 energy of bearing vibration data.

Based on the selected optimal embedding dimension, the SVM model is used to achieve the multisteps prediction. The RBF kernel function is used:

$$K\left(x_i, x_j\right) = \exp\left(-\frac{1}{2\sigma^2}\|x - x_i\|^2\right). \tag{30}$$

In this research the regularity parameter $C$ is set to 90.3, the kernel function parameter $\sigma^2$ is set to 20, and the $\varepsilon$ is set to 0.001. The parameters $C$ and $\varepsilon$ are selected by Particle Swarm Optimization algorithm (PSO) [30]. The popular size

of the PSO is set to 100, the interaction number of the PSO method is set to 20, and the fitness function of the PSO method is set to choose the parameters which make the SVM model fitting error in the training process the smallest. The error goal is set to 0.05, the dimension of the PSO is set to 2, and the inertia weight is set to $w = 0.5$, $c_1 = c_2 = 1.2$. Based on the selected typical features, the features dataset is used to train SVM model and the input features number of SVM is 9 determined by embedding dimension. Then, the trained SVM model is used to predict the bearing running
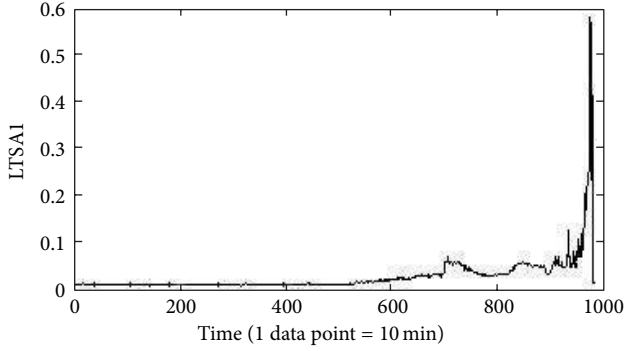
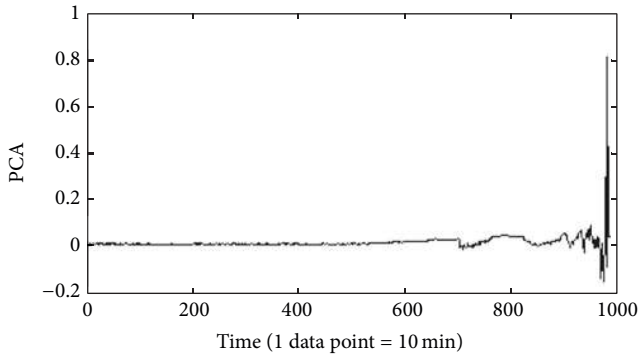Figure 7: The first main projected vector of test bearing based on LTSA method.



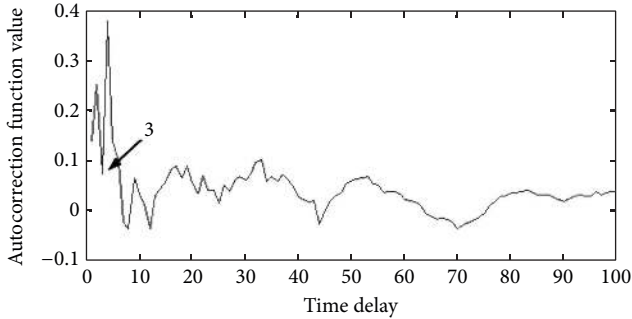Figure 8: The first principal component of test bearing based on PCA method.



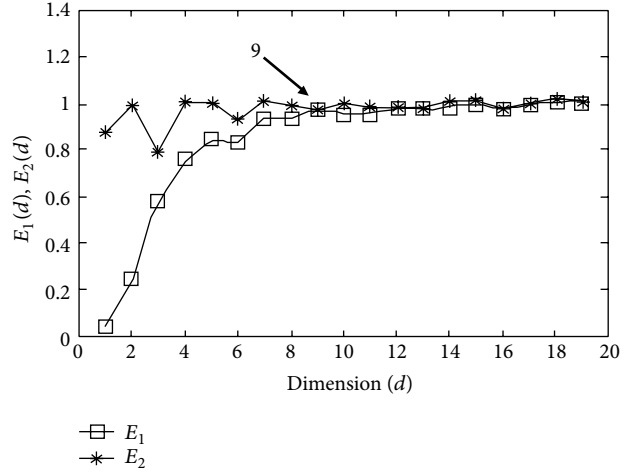Figure 9: Selection of the delay time by the autocorrelation function value of LTSA1.



Figure 10: Selection of the embedding dimension by the CAO method of LTSA1.
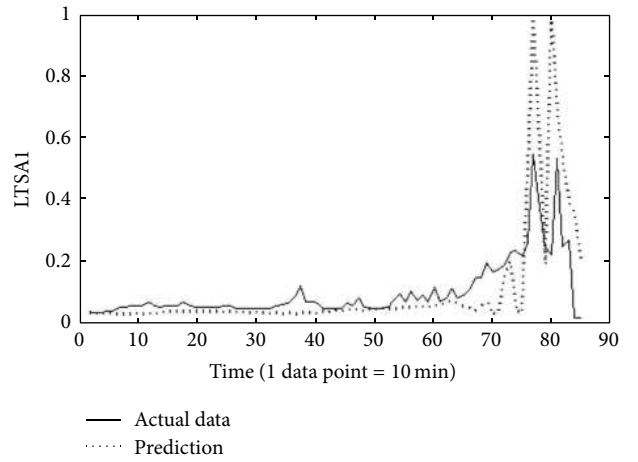


Figure 11: Prediction result based on the SVM model.

state. Before the 700th points, the bearing is working in a normal state, so the 701–900 points are used to train the SVM model and the following 85 points are employed for testing. In order to evaluate the predicting performance, the root-mean square error (RMSE) is utilized as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \widehat{y}_i)^2}{N}}, \tag{31}$$

where $N$ represents the total number of data points in the test set, $y_i$ is actual value in training set or test set, and $\widehat{y}_i$

represents the predicted value of the model. The actual value and the predicted result are shown in Figure 11.

From Figure 11 we can see that the trend of the bearing running state can be predicted by SVM model. From the prediction, we can get a general understanding of the bearing running state in the future, but the predicted result is not accurate, especially the stage of 70–85 point, through calculation. The RMSE of the actual and the predicted result is 0.0469, so the predict result is not satisfied and the prediction error of SVM model (the predicted results subtract the actual data) is shown in Figure 12.

Then the SOM method is used to divide the error into some districts, the iteration number if SOM is 2000; the structure of the state classification matrix is [3×1]. The results of the state division by SOM are shown in Table 2.

It can be seen from Table 2 that when there is downward trend or the point value is less than 0, the state is set to 1 and when there is upward trend, the state is set to 3. According to the classification of the SOM model, the Markov state is divided into the following districts: [−0.10752, −0.00038], [−0.00038, 0.1326], and [0.1326, 0.43612]. The Markov model

TABLE 2: The result of the state division and the prediction results.

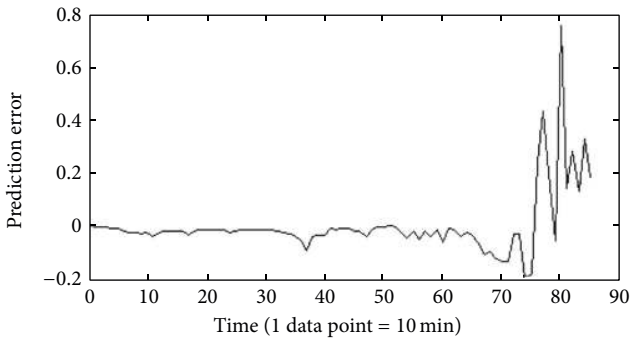| Prediction number | Prediction error | State | Prediction number | Prediction error | State | Prediction number | Prediction error | State |
|---|---|---|---|---|---|---|---|---|
| 1 | −0.001622 | 2 | 31 | −0.014505 | 2 | 61 | −0.009996 | 2 |
| 2 | −0.003811 | 2 | 32 | −0.017872 | 2 | 62 | −0.015532 | 2 |
| 3 | −0.003841 | 2 | 33 | −0.022864 | 2 | 63 | −0.041807 | 1 |
| 4 | −0.007212 | 2 | 34 | −0.031673 | 1 | 64 | −0.022312 | 2 |
| 5 | −0.008162 | 2 | 35 | −0.033233 | 1 | 65 | −0.034537 | 1 |
| 6 | −0.02057 | 2 | 36 | −0.05774 | 1 | 66 | −0.060772 | 1 |
| 7 | −0.023467 | 1 | 37 | −0.090116 | 1 | 67 | −0.10752 | 1 |
| 8 | −0.026518 | 1 | 38 | −0.04199 | 1 | 68 | −0.096034 | 1 |
| 9 | −0.02786 | 1 | 39 | −0.034686 | 1 | 69 | −0.12397 | 1 |
| 10 | −0.027196 | 1 | 40 | −0.032578 | 1 | 70 | −0.13711 | 1 |
| 11 | −0.038575 | 1 | 41 | −0.010655 | 2 | 71 | −0.13712 | 1 |
| 12 | −0.025033 | 1 | 42 | −0.013265 | 2 | 72 | −0.03195 | 1 |
| 13 | −0.018556 | 2 | 43 | −0.006294 | 2 | 73 | −0.02958 | 1 |
| 14 | −0.019786 | 2 | 44 | −0.007407 | 2 | 74 | −0.19645 | 1 |
| 15 | −0.01983 | 2 | 45 | −0.021696 | 2 | 75 | −0.18364 | 1 |
| 16 | −0.020994 | 2 | 46 | −0.017219 | 2 | 76 | 0.24355 | 3 |
| 17 | −0.032989 | 1 | 47 | −0.037728 | 1 | 77 | 0.43612 | 3 |
| 18 | −0.019086 | 2 | 48 | −0.011678 | 2 | 78 | 0.20092 | 3 |
| 19 | −0.013994 | 2 | 49 | −0.001478 | 2 | 79 | −0.05844 | 1 |
| 20 | −0.014597 | 2 | 50 | −0.003989 | 2 | 80 | 0.76262 | 3 |
| 21 | −0.013641 | 2 | 51 | −0.00038 | 2 | 81 | 0.14341 | 3 |
| 22 | −0.014209 | 2 | 52 | −0.00578 | 2 | 82 | 0.28256 | 3 |
| 23 | −0.013578 | 2 | 53 | −0.025056 | 1 | 83 | 0.1326 | 3 |
| 24 | −0.022126 | 2 | 54 | −0.045813 | 1 | 84 | 0.3338 | 3 |
| 25 | −0.019049 | 2 | 55 | −0.01996 | 2 | 85 | 0.18377 | 3 |
| 26 | −0.013825 | 2 | 56 | −0.05286 | 1 | | | |
| 27 | −0.012682 | 2 | 57 | −0.019165 | 2 | | | |
| 28 | −0.014012 | 2 | 58 | −0.039591 | 1 | | | |
| 29 | −0.012089 | 2 | 59 | −0.015265 | 2 | | | |
| 30 | −0.01567 | 2 | 60 | −0.058697 | 1 | | | |



FIGURE 12: The prediction error of the SVM model.

is used to improve the prediction error. For example, the 78 point, the Markov state transition matrix from 1–77 point is

$$p = \begin{bmatrix} 0.697 & 0.2727 & 0.0303 \\ 0.2326 & 0.7674 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \qquad (32)$$

Chose three points 77, 76, and 75 which are recent to the 78 point and set the transfer step as 1, 2, 3; the state prediction results based on the Markov model are shown in Table 3.

From Table 3 we can see that the accumulated value of stage 3 is the largest, so the stage of 78 point is set to 3 and the result is the same with the SOM clustering method.

In this research, according to the stage of the prediction of the Markov model, the correct value is calculated by $\tilde{x}^{(0)}(k) = \hat{x}^{(0)}(k) - \varepsilon$, where $\varepsilon$ is the median value of the divided stage area and $\hat{x}^{(0)}(k)$ is the value predicted though SVM model.

For the 78 point, the corrected value is $0.57298 - (0.1326 + 0.43612)/2 = 0.2956$, where the actual value at this point is 0.37206. Then other points have also been corrected though this method. The corrected results of the point 70–85 are shown in Table 4.

From Table 4 we can see that the Markov model makes the results more precise, which validate the necessary to use the Markov model to improve the effect of the proposed method. The RMSE of the actual and the predicted result is 0.0091, so the prediction accuracy improved significantly.

TABLE 3: Table of the probability status.

| Point | Status probability | Transfer step | State 1 | State 2 | State 3 |
|---|---|---|---|---|---|
| 77 | 3 | 1 | 0 | 0 | 1 |
| 76 | 3 | 2 | 0 | 0 | 1 |
| 75 | 1 | 3 | 0.4757 | 0.4562 | 0.0681 |
| | | Total | 0.4757 | 0.4562 | 2.0681 |

TABLE 4: The corrected results of the points 70–85 based on Markov model.

| Point | SVM model prediction error | Relative error (%) | SVM-Markov model prediction error | Relative error (%) |
|---|---|---|---|---|
| 70 | −0.13711 | 84.37% | −0.1025 | 63.07% |
| 71 | −0.13712 | 79.91% | −0.1235 | 71.97% |
| 72 | −0.03195 | 16.92% | −0.0121 | 6.41% |
| 73 | −0.02958 | 12.85% | 0.0965 | 41.92% |
| 74 | −0.19645 | 84.44% | −0.09512 | 40.88% |
| 75 | −0.18364 | 84.21% | −0.10236 | 46.94% |
| 76 | 0.24355 | 93.60% | 0.01043 | 4.01% |
| 77 | 0.43612 | 79.88% | 0.03542 | 6.49% |
| 78 | 0.20092 | 54.00% | −0.07646 | 20.55% |
| 79 | −0.05844 | 23.60% | −0.01574 | 6.36% |
| 80 | 0.76262 | 344.98% | 0.13262 | 59.99% |
| 81 | 0.14341 | 26.80% | 0.03562 | 6.66% |
| 82 | 0.28256 | 114.19% | 0.20302 | 82.05% |
| 83 | 0.1326 | 49.59% | −0.00251 | 0.94% |
| 84 | 0.3338 | 2061% | 0.03291 | 203.20% |
| 85 | 0.18377 | 1132.15% | 0.06548 | 403.40% |

However, because those points are so far away compared to the predicted point of 1–69, the results still have some error.

In order to compare the predict effect, the most usually used prediction model BP neural networks is used to predict the bearing running state based on the selected features. The learning rate of the neural network and its momentum coefficient are 0.01; the weights are initialized to uniformly distribute random values between −0.1 and 0.1; the iteration number is 2000; the training error is 0.001; the input number is 9; the hidden number is 15; the output node number is 1. The prediction results are shown in Figure 13.

From Figure 13 we can see that the prediction results based on the BPNN model is not working effectively. There are some peaks while in the same position the actual status is not obvious. The RMSE of the predicted result is 0.0932, so the prediction results of the traditional BPNN model is not more effective than the SVM model. In addition, the prediction method based on the BPNN has the problem of prediction results which are unstable; when the same data is used to train and predict, the results are different and even the neural network may fall into the local optimum as shown in Figures 14 and 15.

With the data, the proposed method has also been compared with other methods that had been proposed in relative research. (1) The principal signal features extracted by PCA are utilized by HMM to predict the bearing running state [31]. (2) The time domain and frequency domain features have been directly used as the input of the prediction model, and
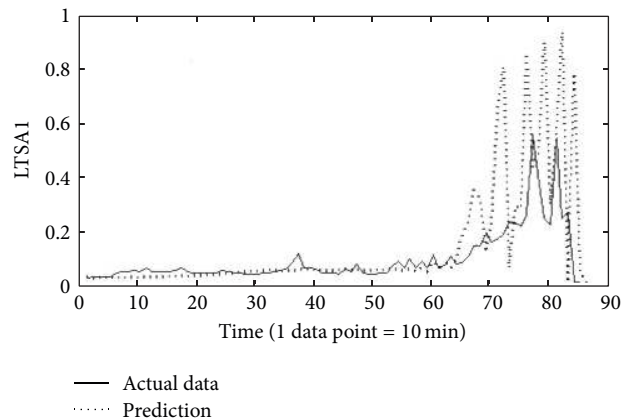


FIGURE 13: Prediction result based on the BPNN model.

the result has been predicted by Neural Network algorithm [1]. (3) The original features have been extracted by PCA as the input of the SVM prediction model [32]. (4) The proposed method in this research. The RMSE of the different methods predicted results is shown in Table 5.

From Table 5, we can see that the RMSE of different prediction methods is very different. The prediction method that the original features have been directly used as the input of the NN model works the worst. This is the reason why the original features are still with high dimension and include superfluous information, which is not appropriate for state

TABLE 5: The RMSE results of different prediction methods.

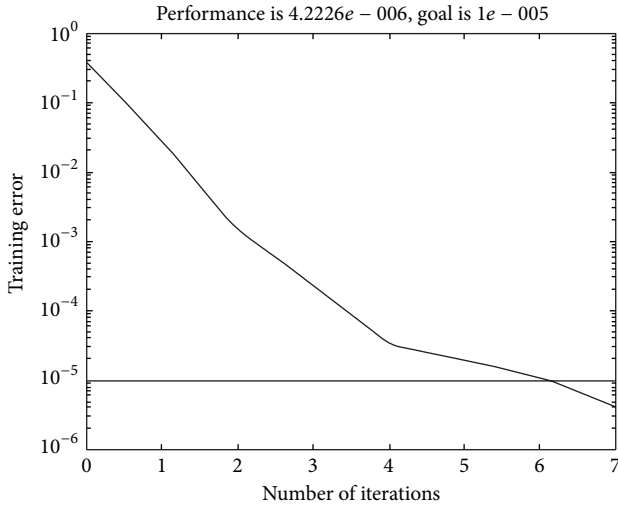| Method | The features extracted by PCA are utilized by HMM prediction model | The features directly used as the input of the Neural Network prediction model | The features extracted by PCA as the input of the SVM prediction model | The proposed method |
|---|---|---|---|---|
| RMSE | 0.0698 | 0.1023 | 0.0521 | 0.0091 |



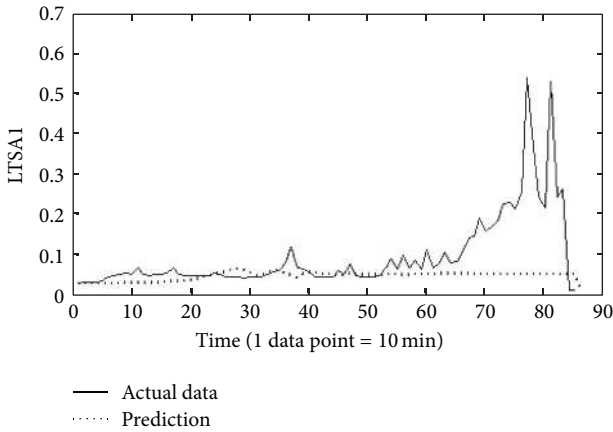FIGURE 14: The training process curve of the BPNN method.



FIGURE 15: The actual data and the result predicted by BPNN fall into the local optimum.

prediction; in addition, the NN prediction model has the drawbacks of slow convergence and difficulty in escaping from local minima. The prediction method based on the HMM model works not more effective than the method based on the SVM model; that is the reason why the HMM is not appropriate for long time forecast. The proposed method works the best; this is because the LTSA features extraction method can effectively extract the typical features and reduce the dimension and the SVM-Markov model can predict the state more precisely than the SVM and Markov model only. So through the comparison we can get that the proposed method is very effective in bearing running state prediction.
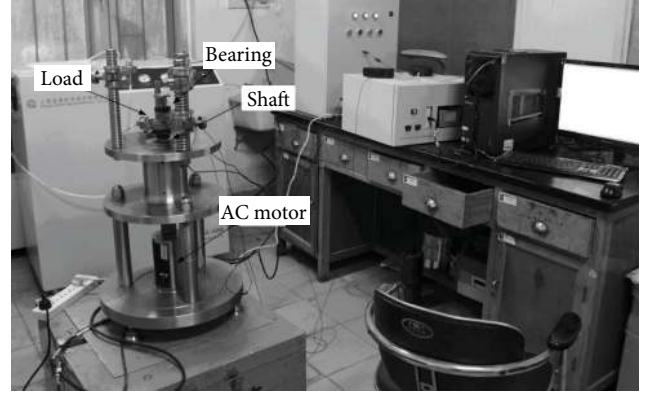


FIGURE 16: Bearing test tig.

5.2. Application. After validating the effectiveness of the proposed method, the method has been used to the actual application. The test rig is shown in Figure 16.

The bearings are hosted on the shaft and the shaft is driven by AC motor. The rotation speed is kept at 1000 rpm and a radial load of 3 kg is added to the bearing. The data sampling rate is 25600 Hz and the data length is 102400 points collected on the date of 2011.11.25 as shown in Figure 17. Every 2 hours, the vibration data are collected for one time. The collected data from 2011.11.25 to 2011.12.17 are analyzed after running for 1 year.

The time domain and time-frequency domain methods are used to deal with the collected vibration data as described in Section 2, Table 1. Then the features are normalized through $\widehat{x} = (x - x_{min})/(x_{max} - x_{min})$ and processed into the interval $[0, 1]$. The LTSA is used to reduce the dimensionality of calculated features and the result is shown in Figure 18.

From Figure 18, we can see that the bearing running state has a fluctuation and upward trend. Especially at 150 points, there is a sudden change of trend, which reflects the bearing's working status change at this moment.

After extracting the typical features, the CAO method is used to determine the embedding dimension of the SVM model. The delay time $\tau$ is set as 2 for the projected vector values though the autocorrelation function as shown in Figure 19.

The embedding dimension is selected by the CAO method. The result is shown in Figure 20 and the optimal embedding dimension $d$ for the projected vector is chosen as 10.

Based on the selected optimal embedding dimension, the SVM model is used to achieve the prediction. The regularity parameter $C$ is set as 909.5 and the $\varepsilon$ is set to 0.01 selected by PSO method. Based on the selected typical features, the features dataset is used to train SVM model, the 1–130 points
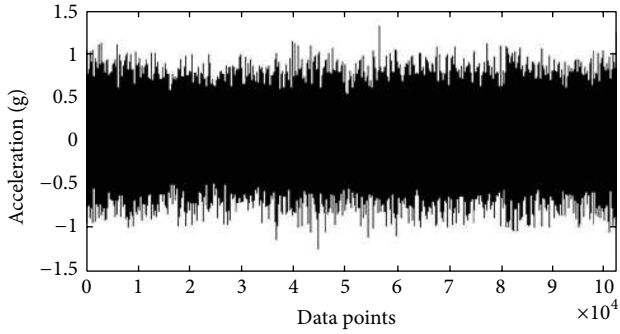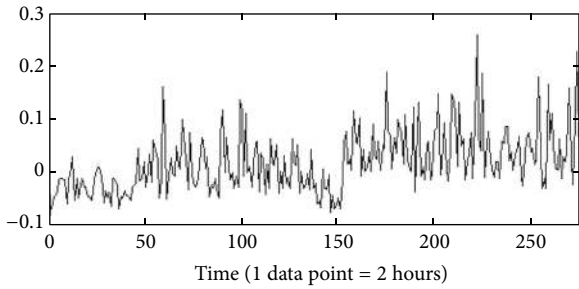
Figure 17: Vibration signal.



Figure 18: The first main projected vector of test bearing based on LTSA method.
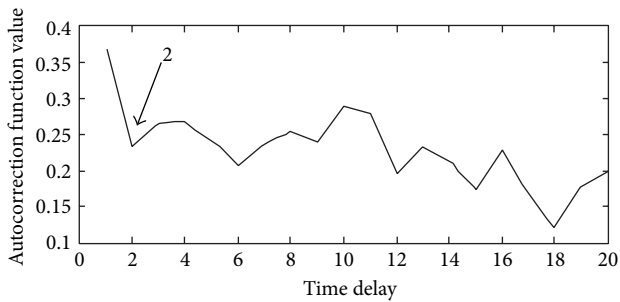


Figure 19: Selection of the delay time by the autocorrelation function value of LTSA1.



Figure 20: Selection of the embedding dimension by the CAO method of LTSA1.
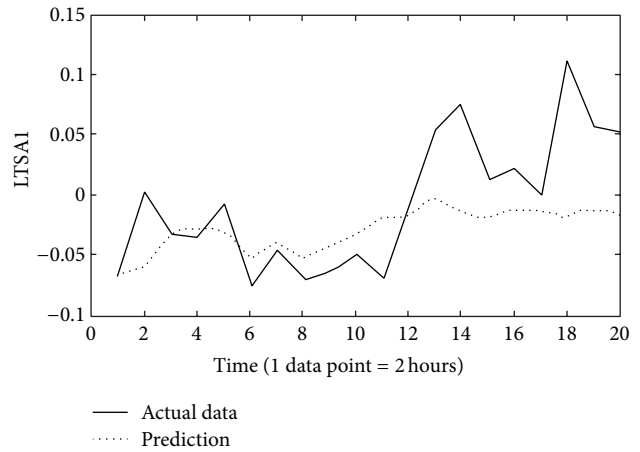


Figure 21: Prediction result based on the SVM model.

are used to train the SVM model, and the following 20 points are employed for testing. The prediction results are shown in Figure 21.

From Figure 21 we can see that the trend of the bearing running state can be predicted by SVM model; from the prediction, we can get a general upward trend similar to the actual status. In addition, the sudden change of points 8 to 12 (near the 150 points in original signal as mentioned in Figure 17) is also showed out. However, the results are still not precise. In order to improve the prediction effect, the SOM method is used to divided the prediction error into some districts, the iteration number if SOM is 1000; the structure of the state classification matrix is [3 × 1]. The results of the state division by SOM are shown in Table 6.

Based on the classification of the SOM model, the Markov state is divided into the following districts [−0.13365, −0.034682], [−0.034682, 0], and [0, 0.053935];
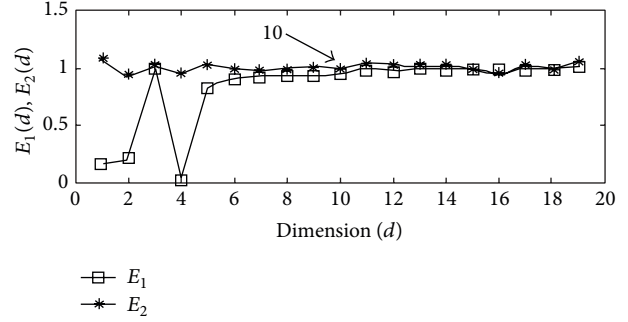
the Markov model is used to improve the prediction error. The corrected results of the points 1–20 are shown in Table 7.

From Table 7, we can see that the Markov model make the results more precise.

Through the validation and actual application result we can see that the proposed method can predict the future status of the bearing, which is necessary for us to make some plan and do maintenance to reduce the risk of unnecessary accident.

## 6. Conclusions

(1) The time domain and time-frequency domain methods are used to extract the original features from the mass vibration data, and in order to reduce the original features dimension and the superfluous information of the original features, the multifeatures fusion technique LTSA is used to fusion the original features and reduce the dimension.

(2) Use the proposed SVM model to achieve bearing running state prediction. The proposed approach is validated by real-world vibration signals. The results show that the proposed methodology is of high accuracy, which is effective for the bearing running state prediction.

TABLE 6: The result of the state division and the prediction results.

| Point | Prediction error | State | Point | Prediction error | State |
|-------|------------------|-------|-------|------------------|-------|
| 1 | 0 | 3 | 11 | 0.053935 | 3 |
| 2 | −0.063821 | 1 | 12 | −0.0052358 | 1 |
| 3 | 0.002906 | 3 | 13 | −0.058422 | 1 |
| 4 | 0.0074271 | 3 | 14 | −0.094039 | 1 |
| 5 | −0.021835 | 2 | 15 | −0.032173 | 2 |
| 6 | 0.024138 | 3 | 16 | −0.034682 | 2 |
| 7 | 0.0059543 | 3 | 17 | −0.014101 | 2 |
| 8 | 0.016244 | 3 | 18 | −0.13365 | 1 |
| 9 | 0.022316 | 3 | 19 | −0.06811 | 1 |
| 10 | 0.016692 | 3 | 20 | −0.070675 | 1 |

TABLE 7: The corrected results of the points 1–20 based on Markov model.

| Point | SVM model prediction error | Relative error (%) | SVM-Markov model prediction error | Relative error (%) |
|-------|----------------------------|--------------------|-----------------------------------|--------------------|
| 1 | 0 | 0 | 0 | 0 |
| 2 | −0.0103821 | 579.21% | −0.006361 | 178.31% |
| 3 | 0.002906 | 8.5739% | 0.000603 | 1.7791% |
| 4 | 0.0074271 | 20.311% | −0.001021 | 2.7922% |
| 5 | −0.021835 | 274.63% | −0.009362 | 117.75% |
| 6 | 0.024138 | 30.429% | 0.013254 | 16.708% |
| 7 | 0.0059543 | 12.666% | 0.002615 | 5.5628% |
| 8 | 0.016244 | 22.459% | 0.010027 | 13.863% |
| 9 | 0.022316 | 33.565% | 0.009565 | 14.386% |
| 10 | 0.016692 | 32.417% | 0.009473 | 18.397% |
| 11 | 0.053935 | 73.825% | 0.021563 | 29.515% |
| 12 | −0.0052358 | 34.947% | −0.002355 | 15.719% |
| 13 | −0.058422 | 107.06% | −0.019347 | 35.455% |
| 14 | −0.094039 | 122.79% | −0.014236 | 18.589% |
| 15 | −0.032173 | 263.1% | −0.009878 | 80.779% |
| 16 | −0.034682 | 159.3% | −0.008969 | 41.197% |
| 17 | −0.004101 | 355.3% | −0.001892 | 199.5% |
| 18 | −0.13365 | 117.18% | −0.050841 | 44.576% |
| 19 | −0.06811 | 120.92% | −0.028323 | 50.282% |
| 20 | −0.070675 | 135.04% | −0.014267 | 27.152% |

(3) This research gives an example of combined approaches for the bearing running state prediction. Through analysis and validation we can get that the proposed method takes good use of the advantages of each part and achieve a high recognition accuracy and efficiency.

(4) As the redundancy increases, the complexity of computation increases as well. This is one of the main shortcomings of the proposed method, which will be explored in the future.

## Conflict of Interests

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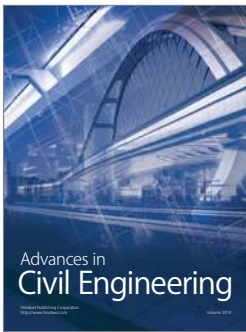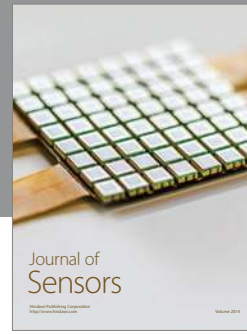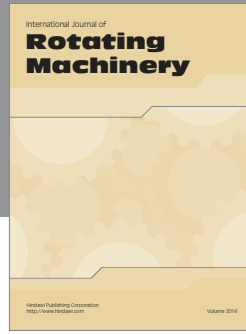
## Acknowledgments

## References

[1] P. L. Zhang, B. Li, and S. S. Mi, "Bearing fault detection using multi-scale fractal dimensions based on morphological covers," *Shock and Vibration*, vol. 19, no. 6, pp. 1373–1383, 2012.

[2] S. Dong and T. Luo, "Bearing degradation process prediction based on the PCA and optimized LS-SVM model," *Measurement*, vol. 46, pp. 3143–3152, 2013.

[3] L. L. Jiang, Y. L. Liu, X. J. Li, and A. Chen, "Degradation assessment and fault diagnosis for roller bearing based on AR model and fuzzy cluster analysis," *Shock and Vibration*, vol. 18, no. 1-2, pp. 127–137, 2011.

[4] J. B. Yu, "Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models," *Mechanical Systems and Signal Processing*, vol. 25, no. 7, pp. 2573–2588, 2011.

[5] J. H. Yan, C. Z. Guo, and X. Wang, "A dynamic multi-scale Markov model based methodology for remaining life prediction," *Mechanical Systems and Signal Processing*, vol. 25, no. 4, pp. 1364–1376, 2011.

[6] H. Ocak, K. A. Loparo, and F. M. Discenzo, "Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: a method for bearing prognostics," *Journal of Sound and Vibration*, vol. 302, no. 4-5, pp. 951–961, 2007.

[7] C. Sun, Z. S. Zhang, and Z. J. He, "Research on bearing life prediction based on support vector machine and its application," *Journal of Physics*, vol. 305, no. 1, Article ID 012028, 2011.

[8] J. Wang, G. H. Xu, Q. Zhang, and L. Liang, "Application of improved morphological filter to the extraction of impulsive attenuation signals," *Mechanical Systems and Signal Processing*, vol. 23, no. 1, pp. 236–245, 2009.

[9] Y. B. Zhan and J. P. Yin, "Robust local tangent space alignment via iterative weighted PCA," *Neurocomputing*, vol. 74, no. 11, pp. 1985–1993, 2011.

[10] J. Wang, W. Jiang, and J. Gou, "Extended local tangent space alignment for classification," *Neurocomputing*, vol. 77, no. 1, pp. 261–266, 2012.

[11] S. Dong, B. Tang, and R. Chen, "Bearing running state recognition based on non-extensive wavelet feature scale entropy and support vector machine," *Measurement*, vol. 46, no. 10, pp. 4189–4199, 2013.

[12] Y. Qin, B. P. Tang, and J. X. Wang, "Higher-density dyadic wavelet transform and its application," *Mechanical Systems and Signal Processing*, vol. 24, no. 3, pp. 823–834, 2010.

[13] A. Moosavian, H. Ahmadi, and A. Tabatabaeefar, "Comparison of two classifiers; K-nearest neighbor and artificial neural network, for fault diagnosis on a main engine journal-bearing," *Shock and Vibration*, vol. 20, no. 2, pp. 263–272, 2013.

[14] T. Ghidini and C. Dalle Donne, "Fatigue life predictions using fracture mechanics methods," *Engineering Fracture Mechanics*, vol. 76, no. 1, pp. 134–148, 2009.

[15] S. Marble and B. P. Morton, "Predicting the remaining life of propulsion system bearings," in *Proceedings of the 2006 IEEE Aerospace Conference*, Big Sky, Mont, USA, March 2006.

[16] Z. G. Tian, L. N. Wong, and N. M. Safaei, "A neural network approach for remaining useful life prediction utilizing both failure and suspension histories," *Mechanical Systems and Signal Processing*, vol. 24, no. 5, pp. 1542–1555, 2010.

[17] W. Caesarendra, A. Widodo, and B. Yang, "Combination of probability approach and support vector machine towards machine health prognostics," *Probabilistic Engineering Mechanics*, vol. 26, no. 2, pp. 165–173, 2011.

[18] J. Lee, J. Ni, D. Djurdjanovic, H. Qiu, and H. Liao, "Intelligent prognostics tools and e-maintenance," *Computers in Industry*, vol. 57, no. 6, pp. 476–489, 2006.

[19] R. Q. Huang, L. F. Xi, X. L. Li, C. Richard Liu, H. Qiu, and J. Lee, "Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods," *Mechanical Systems and Signal Processing*, vol. 21, no. 1, pp. 193–207, 2007.

[20] C. W. Fei and G. C. Bai, "Wavelet correlation feature scale entropy and fuzzy support vector machine approach for aero-engine whole-body vibration fault diagnosis," *Shock and Vibration*, vol. 20, no. 2, pp. 341–349, 2013.

[21] P. C. Gonçalves, A. R. Fioravanti, and J. C. Geromel, "Markov jump linear systems and filtering through network transmitted measurements," *Signal Processing*, vol. 90, no. 10, pp. 2842–2850, 2010.

[22] A. N. Jiang, S. Y. Wang, and S. L. Tang, "Feedback analysis of tunnel construction using a hybrid arithmetic based on Support Vector Machine and Particle Swarm Optimisation," *Automation in Construction*, vol. 20, no. 4, pp. 482–489, 2011.

[23] V. T. Tran, B. Yang, and A. C. C. Tan, "Multi-step ahead direct prediction for the machine condition prognosis using regression trees and neuro-fuzzy systems," *Expert Systems with Applications*, vol. 36, no. 5, pp. 9378–9387, 2009.

[24] M. Marcellino, J. H. Stock, and M. W. Watson, "A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series," *Journal of Econometrics*, vol. 135, no. 1-2, pp. 499–526, 2006.

[25] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D*, vol. 110, no. 1-2, pp. 43–50, 1997.

[26] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence*, D. A. Rand and L. S. Young, Eds., pp. 366–381, Springer, New York, NY, USA, 1981.

[27] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, no. 2, pp. 1134–1140, 1986.

[28] T. Kohonen, *Self-Organizing Maps*, Springer, Berlin, Germany, 1995.

[29] J. Lee, H. Qiu, G. Yu, and J. Lin, "Rexnord Technical Services, "Bearing Data Set", IMS, University of Cincinnati, NASA Ames Prognostics Data Repository," NASA Ames, Moffett Field, CA, http://ti.arc.nasa.gov/project/prognostic-data-repository/ .

[30] H. Mohkami, R. Hooshmand, and A. Khodabakhshian, "Fuzzy optimal placement of capacitors in the presence of nonlinear loads in unbalanced distribution networks using BF-PSO algorithm," *Applied Soft Computing Journal*, vol. 11, no. 4, pp. 3634–3642, 2011.

[31] X. D. Zhang, R. G. Xu, C. Kwan, S. Y. Liang, Q. Xie, and L. Haynes, "An integrated approach to bearing fault diagnostics and prognostics," in *Proceedings of the American Control Conference (ACC '05)*, pp. 2750–2755, Portland, Ore, USA, June 2005.

[32] C. Sun, Z. S. Zhang, and Z. J. He, "Research on bearing life prediction based on support vector machine and its application," *Journal of Physics*, vol. 305, no. 1, Article ID 012028, pp. 1–9, 2011.

Journal of
Engineering

The Scientific
World Journal

International Journal of
Rotating
Machinery

Journal of
Sensors

International Journal of
Distributed
Sensor Networks

Advances in
Civil Engineering

Journal of
Control Science
and Engineering

Journal of
Robotics

Journal of
Electrical and Computer
Engineering

Advances in
OptoElectronics

VLSI Design

International Journal of
Navigation and
Observation

Modelling &
Simulation
in Engineering

International Journal of
Aerospace
Engineering

International Journal of
Chemical Engineering

International Journal of
Antennas and
Propagation

Active and Passive
Electronic Components

Shock and Vibration

Advances in
Acoustics and Vibration

Hindawi

Submit your manuscripts at
http://www.hindawi.com