This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2943507, IEEE Transactions on Industrial Informatics

Intelligent Diagnostics for Bearing faults Based on Integrated Interaction of Nonlinear Features

Lin Bo, Xiaofeng Liu, Guanji Xu

***Abstract*—An unforeseen fault of the key bearing of production system due to different reasons has the potential to cause an interruption in the entire production line, resulting in economic and production losses. To improve the reliability of industry production, this paper presents an intelligent diagnosis method for element rolling bearing based on the integrated interaction relationship of vibration nonlinear features. The nonlinear features of vibration signals are extracted using recurrence quantification analysis (RQA) and regrouped into different subsets of non-redundant features with the same level of discrimination ability through the technology of ReliefF-affinity propagation clustering. The weighted voting variable predictive model class discrimination (WV-VPMCD) is proposed to fully utilize the interaction of RQA features to do intelligent diagnostics for bearing faults. The experimental results showed that the WV-VPMCD outperformed the conventional intelligent diagnosis methods in terms of accuracy, consistency, stability and robustness, especially in the case of small number of samples.**

***Index Terms*—Bearing fault, feature interaction, intelligent diagnostics, nonlinear features, variable predictive model class discrimination.**

I. INTRODUCTION

Rotating machines are widely used in the manufacturing industry and usually operate by means of rolling bearings that may be affected by several faults which cause machine breakdown and performance level reduction [1]. Effective diagnosis allows users to plan maintenance to fix or remove the affected bearing, thus reducing system downtime and increasing safety and reliability [2]. Aiming at automated detection and diagnosis of machine conditions, intelligent fault diagnosis methods employ artificial intelligence (AI) techniques and extract the underlying knowledge from the historic sensor data [3,4]. Technically, combining with the feature extraction approaches, they can effectively process massive data and largely increase the reliability of fault detection and diagnosis system.

Manuscript received January 10,2019; accepted September 9, 2019. Data of publication October, 11, 2019; data of current version June 1, 2019. This work was supported in part by the National Science Foundation of China under Grand NSFC-51975067, 51675064; Paper no. TII-19-1512 (Corresponding author: Xiaofeng Liu)

The authors are with the Department of Mechanical Engineering, Chongqing University, Chongqing, China (e-mail:bolin0001@aliyun.com; liuxfeng0080@126.com; 344533315@qq.com)

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Intelligent diagnosis for bearing faults is an extremely challenging task because the bearing in a fault state may behave complex dynamical uncertainty and its vibration signal usually is strongly non-linear and non-stationary. A number of non-linear parameter estimation methods and theories, such as such as fractal dimension [5], wavelet entropy [6] and sample entropy [7], waveform singularity exponent [8] etc. have been introduced to bearing fault diagnosis. However, these methods generally depend heavily on the data length and lack enough robustness to noise interferences. A useful tool to deeply investigate the nonlinear dynamics is recurrence quantification analysis (RQA), which is a nowadays-recognized good method to deal with non-stationary, non-linear and relatively short data series [9]. RQA has an increasing number of applications in processing nonlinear chaotic behaviors of manufacturing system, such as investigating nonstationary work in the diesel engine acceleration [10], studying car vibration transient behavior [11] and identifying rotor-bearing system resonance states [12] etc. In this paper, RQA is utilized to extract the defect-related features hidden in the collected vibration signal of rolling bearing.

To realize the automatic identification of production system states, some AI technologies including artificial neural networks (ANN) [13], Support Vector Machine (SVM) [2], convolutional neural network [14], hidden semi-Markov model [15] etc. have been successfully employed. It is noteworthy that the performance of these conventional methods depends on choosing appropriate parameters obtained from the big training samples. The variable prediction model based class discrimination (VPMCD) [16] is a new AI technique targeting the mutual interrelationship between feature values, which has no preset parameters and avoids the iterative process of conventional neural networks and the parameter optimization of general recognition models. It meets the challenge of pattern recognition when different classes cannot be distinguished just based on decision boundaries or conditional discriminating rules.

However, for a complicated system with high-dimension feature vectors, it is hard to determine the natural relationship of features interaction for characterizing the system. Hence, it is quite necessary to select the relevant features that possess better discrimination ability and less redundancy. Hence, ReliefF ranking and affinity propagation (AP) clustering are introduced to obtain some compacted feature subsets for improving the generalization capability and identification accuracy of VPMCD. In addition, when there are some outliers or local

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2943507, IEEE Transactions on Industrial Informatics

deviations in the feature dataset, the established variable predictive model (VPM) will deviate from the real model. Aiming to these problems discussed above, some optimization algorithms and kernel functions have been introduced into the improved methods, such as robust regression-VPMCD [17], neuro network VPMCD [18], radial basis Function-VPMCD [19], genetic algorithm VPMCD [20], Kernel partial least squares regression based VPMCD [21], etc. Because of the iteration or tuning of parameters, these improved methods have actually deviated from the essence of VPMCD, thus suffer from being trapped in local optima and high computational cost. Moreover, these methods generally need enough and regular training samples to get the accurate models. In the case of small or singular samples, their poor generalization abilities lead to inaccurate identification results.

To meet these difficulties, combining weighted voting based on Bagging set ensemble algorithm and VPMCD, the weighted voting VPMCD (WV-VPMCD) is proposed in this paper. The primary novelty of this paper is such an ensemble intelligent diagnosis method, which integrates the mutual interaction of all the features and considers the uncertainty of selecting training samples. It overcomes the perturbation caused by different samples distribution and the instability of different classifiers, thus has high identification accuracy and good stability even under the condition of small samples and high-dimension feature eigenvectors.

II. RECURRENCE QANTIFICATION ANALYSIS

RQA is a method of nonlinear data analysis for the investigation of dynamical system, aiming to quantify differently appearing recurrence plot (RP) based on small-scale structures. RP is a graphical tool that visualizes the recurrence behavior of the phase space trajectory of dynamical system. For a series $\{s_1, s_2,\ldots, s_N\}$with length N, it is embedded into the space $\mathbf{R}^m$ with embedding dimension m and time delay τ according to nonlinear dynamic theory. The RP can be briefly described as

$$R_{i,j} = \Theta\left(\varepsilon - \left\|s_i - s_j\right\|\right),\ i, j = 1,\cdots, N \quad (1)$$

where ||.|| and Θ(.) is the norm and the Heaviside function, respectively. This means if the distance between s_i and s_j is less than ε, then $R_{i,j}$=1 and a dot is placed at (i, j) in the RP. The scale ε is a cutoff distance defining a sphere centered at s_i. Since RP just is a visualization tool, low screen and printer solutions can worsen the interpretation of its pattern and structure, so RQA is used to quantify the number and duration of small-scale structures within the RP by utilizing several calculated recurrence statistics. The core of RQA is the computation of several statistics providing the identification and the quantification of transient recurrent patterns characterizing the dynamic change behavior of the time series, which is based on the recurrence point density and the diagonal and vertical line structures of RPs [9], [12]. In this paper, 11 RQA features(listed in Table I) are used to characterize the bearing vibration signals collected in different fault states.

TABLE I
RQA FEATURES

ID	Feature parameter	Equation
1	Recurrence Rate(RR)	$RR = \frac{1}{N^2}\sum_{i,j=1}^{N} R_{i,j}(\varepsilon)$
2	Determinism DET	$DET = \sum_{l=l_{\min}}^{N} lP^{\varepsilon}(l) \Big/ \sum_{i,j=1}^{N} R_{i,j}^{m,\varepsilon}$
3	Mean diagonal line length L	$DET = \sum_{l=l_{\min}}^{N} lp(l) \Big/ \sum_{l=l_{\min}}^{N} p(l)$
4	Maximal diagonal line length $L_{\max}$	$L_{\max} = \max\left(\{l_i, i = 1,2,\cdots,\ N_l\}\right)$
5	Entropy ENTE	$ENTR = -\sum_{l=l_{\min}}^{N} p(l)\ln p(l)$
6	Laminarity LAM	$LAM = \sum_{v=v_{\min}}^{N} vp^{\varepsilon}(v) \Big/ \sum_{v=1}^{N} vp^{\varepsilon}(v)$
7	Trapping Time TT	$TT = \sum_{v=v_{\min}}^{N} vp^{\varepsilon}(v) \Big/ \sum_{v=v_{\min}}^{N} p^{\varepsilon}(v)$
8	Maximal vertical line length $V_{\max}$	$V_{\max} = \max\left(\{v_i, i = 1,2,\cdots,\ N_v\}\right)$
9	Recurrence Time of 1st type T_j^1	$T_j^1 = \left\|\left\{i, j : \vec{s}_i, \vec{s}_j \in R_i\right\}\right\|$
10	Recurrence Time of 2nd type T_j^2	$T_j^2 = \left\|\left\{i, j : \vec{s}_i, \vec{s}_j \in R_i; \vec{s}_{j-1} \notin R_i\right\}\right\|$
11	Recurrence Time Entropy (RET)	$RET = -p(TT)\ln(TT)$

Notes：$P^{\varepsilon}(l)=\{l_i, i=1,2,\ldots,N_l\}$ is the frequency distribution of the length l of diagonal structure; N_l is the absolute number of diagonal lines; $P^{\varepsilon}(v)=\{v_i, i=1,2,\ldots,N_v\}$ is the frequency distribution of the length l of vertical structure; $p(l)= P^{\varepsilon}(l)/\sum_{l=lmin}^{N} P^{\varepsilon}(l)$ is the probability that a diagonal line has exactly length l. R_i denotes the recurrence points which belong to the trajectory $\vec{s}_i$.

III. MULTI-CLASSFICATION OF VPMCD

For the pattern recognition issue with L different failure categories, d different feature values $\mathbf{x}=[x_1, x_2, \ldots, x_d]$ from n signal vibrations are used to describe each kind of failure, thus a training set $N[n_l \times d; L]$ with n_l being number of samples belonging to class $l(l=1, 2,\cdots, L)$. The basis of VPMCD is that the interaction relationship between x_i and $x_{j\neq i}$ (i=1,2,…,d) is different in different failure category. The structure of inter-feature variable associations for each failure category is established using VPM which uniquely characterize the variable associations for that class and vary between different failure categories. That is, for any feature variable x_i^l (i=1,2,…,d) can be predicted by polynomial about $x_{j\neq i}$ in the model VPM_i^l. Generally, there exist four polynomial model types of VPM_i^l [21], as shown in (2) to (5):

Linear VPM:

$$x_i = b_0 + \sum_{j=1}^{r} b_j x_j \quad (2)$$

Line and interaction VPM:

$$x_i = b_0 + \sum_{j=1}^{r} b_j x_j + \sum_{j=1}^{r}\sum_{k=j+1, k\neq i}^{r} b_{jk} x_j x_k \quad (3)$$

Quadratic VPM:

$$x_i = b_0 + \sum_{j=1}^{r} b_j x_j + \sum_{j=1}^{r} b_{jj} x_j^2 \quad (4)$$

Quadratic and interaction VMP:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2943507, IEEE Transactions on Industrial Informatics

$$x_i = b_0 + \sum_{j=1}^{r} b_j x_j + \sum_{j=1}^{r} b_{jj} x_j^2 \sum_{j=1}^{r} \sum_{k=j+1, k\neq i}^{r} b_{jk} x_j x_k \tag{5}$$

where $r(r \le d\text{-}1)$ is the predictor order; b_0, b_j, b_{jj} and b_{jk} are model parameters of VPM_i^l. For any testing sample with feature x_i, the established VPM_i^l is used to obtain the predict value $x_{i\,pred}^l$ which has a higher likelihood of being similar as x_i if the sample belongs to class l. The sum of squared prediction error $\sum_{i^p=1}(x_i\text{-}x_{i\,pred}^l)^2$ of all feature values are calculated for different failure category. Based on the minimum squared predictor error, the testing sample can be considered to belong to class l.

There exist three problems to be solved in the application of VPMCD. Firstly, the higher the feature dimension, the more training samples are required to establish the VPM. It means that small-sample data and high-dimensional eigenvector may provide lower statistical strength to parameters estimated during the training process. Secondly, since the VPMCD heavily depends on the complex interaction amongst the features, the irrelevant and redundant features can greatly degrade the performance of VPMCD. Finally, if there are too many outliers in the training dataset, the established VPM will deviate from the inherent interrelationship of features. The key to solve these problems is to fully utilize the intrinsic relationship between non-redundant features and maximize the assistance of existing VPMs. This is especially important for the feature set with overlapping patterns that cannot be efficiently separated based on distance measures or decision boundaries.

IV. FEATURE SUBSET SELECTIONU

The main goal of the feature subset selection is to improve the generalization capability of VPMCD by compacting the feature space, eliminating redundant features and selecting the most informative ones. We proposed a new feature selection method for generating a candidate subset of features, inspired by ReliefF score ranking and AP clustering, so the new method is named RSAP. The ReliefF assigns a ranking score for each feature individually based on k nearest neighbor algorithm [22]. The ReliefF score (RS) for feature x_k is given as

$$\begin{aligned} RS(x_k) = {} & P(\text{different value of } x_k \mid \text{different class}) \\ & - P(\text{different value of } x_k \mid \text{same class}) \end{aligned} \tag{6}$$

$$NRS = RS(x_k) / \sum_{k=1}^{d} RS(x_k) \tag{7}$$

where P means probability. The RS is normalized to get the NRS (7) of kth feature between (0, 1), where the biggest value of NRS means kth feature is ranked top1 (the best) and the smallest one means the worst.

Affinity propagation (AP) is an exemplar-based clustering algorithm, which takes the input similarities between features and finds clusters with small error and fast execution speed [23]. To determine the final clustering result, two kinds of information between two features are defined as responsibility $r(i,k)$ and availability $a(i,k)$. The $r(i,k)$ sent from feature i to feature k reflects the appropriate degree of feature k as the group center of feature i, as shown in (8).

$$r(i,k) \leftarrow s(i,k) - \max_{k' \neq k} \{a(i,k') + s(i,k')\} \tag{8}$$

where $s(i,k)$ is the value of similarity between each feature's pair$\{x_i, x_k\}$. The $a(i,k)$ sent from the candidate group center feature k to the feature i, indicates the evidence of how appropriate it would be to the point i to choose the point k as an exemplar. It is described as

$$\begin{cases} a(i,k) \leftarrow min\left\{0, r(k,k) + \sum_{i' \notin \{i,k\}} max\{0, r(i',k)\}\right\} \\ a(k,k) \leftarrow max\{0, r(i',k)\} \end{cases} \tag{9}$$

In the process of clustering, the $r(i,k)$ and $a(i,k)$ are started with diagonal elements of similarity matrix and a null vector, respectively. Thus, the definition of an exemplar occurs when the combination of maximum value:

$$AP = \max\{r(i,k) + a(i,k)\} \tag{10}$$

The main steps of RSAP are described as

1) Get the NRS of each feature according to (6)-(7) and order the all features in descending order of their rank.
2) Cluster the all features with AP algorithm in (8)-(10), then get a number of clusters of features
3) Order the features in each cluster in decreasing rank of their NRS values
4) Group, all first order features in each cluster into the 1st feature subset, 2nd order features into the second subset and so on
5) If the number of features in each cluster are not equal, then remove the last feature from the cluster that has more features.

Based on the above steps, iteration for multiple times, the different feature subsets can be obtained. In each subset, the features have less similarity, more independence and lower dimension, which are beneficial to establish a precise VPM. The VPMs corresponding to different feature subsets are complementary in the final decision to a certain degree.

V. WEIGHTED VOTING-VPMCD

Classification accuracy of VPM_k is often calculated by referring to labels of sampled data instances (used in training). However, the performance that VPM_k achieves on the dataset used in its training may not be an ideal indicator to its performance on new and unseen data instances. That is, assigning weights to the member VPMCDs according to how they perform in training may increase the risk of overestimating what they will achieve. In this paper, the risk is reduced by using un-sampled data samples (not used in training) to evaluate and weight member VPMCDs. In order to improve the accuracy of VPMCD and maintain its stability and consistency,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2943507, IEEE Transactions on Industrial Informatics

a novel intelligent diagnosis method named WV-VPMCD is proposed. In this method, the training dataset is regrouped into several feature subsets with the RSAP method, and then an integrated model is constructed by the sub-VPMs adaptively according to weighted bagging ensemble algorithm. The proposed method is illustrated in Fig.1 and the detailed steps are provided as follows:

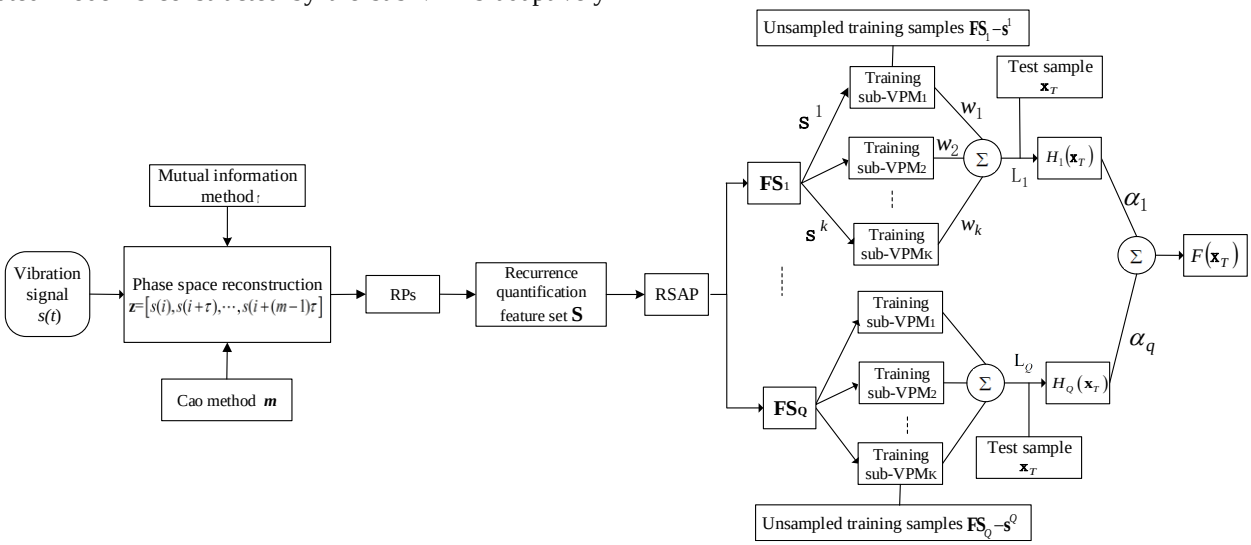


Fig. 1 Flow chart of WV-VPMCD

1) With the confirmed values of m and τ, the RP of vibration signal $s(i)$ can be obtained based on the phase reconstruction $\mathbf{z}=[s(i), s(i+\tau),\ldots,s(i+(m-1)\tau]$. The choice of m is usually based on counting false nearest neighbors when increasing m, and choosing the value of m where the number of false nearest neighbors goes to zero [24]. The delay τ must be selected to minimize the autocorrelations between the time series points. In practice, 'reasonable' value of τ corresponds to the first minimum of the mutual information function [25].

2) To quantify differently appearing RPs, the RQAs of each signal are calculated and noted as $\mathbf{x}=[RR, DET, L, L_{max}, ENTE, LAM, TT, V_{max}, T_j^1, T_j^2, RET]$. So, the training sample set $\mathbf{S}=\{(\mathbf{x}_n,y_n)|\ n=1,\ldots,N)\ \}$ with $N\times d$ (N is the number of training feature samples and d=11) is formed, where $y_n\in\{\ l_1, l_2,\ldots, l_C\}$ is the class label of $\mathbf{x}_n$.

3) With the method RSAP introduced in Section IV, the 11 features of $\mathbf{x}$ is ranked and $\mathbf{S}$ is partitioned into P feature clusters$\{\mathbf{FC_1}, \mathbf{FC_2},\ldots, \mathbf{FC_P}\}$, and then the new feature subsets $\{\mathbf{FS_1}, \mathbf{FS_2},\ldots, \mathbf{FS_Q}\}$ are generated.

4) Generate a series of "bagged" sets $\mathbf{s}^k$ by resampling with replacement from each feature subset $\mathbf{FS}_q(q=1,2, \ldots,Q)$. Each $\mathbf{s}^k$ (k=1,2,…,K) is used to train the member VMPs, and then get ensemble classifier $\mathrm{L}_q=\{\mathrm{VPM}_1,\ldots,\mathrm{VPM}_K\}$.

5) Use the un-sampled feature set (noted as $\mathbf{FS}_q$- $\mathbf{s}^k$) in $\mathbf{FS}_q$ to test VPM_k and get the discrimination rate p_k that the un-sampled training instances are correctly classified by VPM_k, the weighted coefficient of VPM_k is set as

$$w_k=\frac{\log\left[p_k/(1-p_k)\right]}{\sum_{k=1}^{K}\log\left[p_k/(1-p_k)\right]} \tag{11}$$

6) For the $\mathbf{FS}_q$, the general form of combining individual VPM classifiers is given as follows:

$$H_q(\mathbf{x}_T)=\max_{l_c,c=1,\cdots,C}\left(\sum_{i=1}^{K} g_{kc}(\mathbf{x}_T)\cdot w_k\right) \tag{12}$$

where $H_q(\mathbf{x}_T)$ represents the predictive label of an unlabeled sample $\mathbf{x}_T$ based on the $\mathbf{FS}_q$, $g_{kc}(\mathbf{x}_T)$ is the probability of $\mathbf{x}_T$ classified to l_c by VPM_k.

7) Calculate the total training accuracy of L_q for the $\mathbf{FS}_q$, i.e. $ACC_q=\sum_{k=1}^{K}p_k$, and the corresponding weight of L_q is set as

$$\alpha_q=\frac{ACC_q}{\sum_{q=1}^{Q}ACC_q}. \tag{13}$$

8) The final prediction result is

$$F(\mathbf{x}_T)=\max_{l_c,c=1,\cdots,C}\left(\sum_{q=1}^{Q}H_{qc}(\mathbf{x}_T)\cdot a_q\right) \tag{14}$$

where $F(\mathbf{x}_T)$ is the final label of an unlabeled sample $\mathbf{x}_T$, $H_{qc}(\mathbf{x}_T)$ is the probability of $\mathbf{x}_T$ classified to l_c by L_q .

The basic idea of the WV-VPMCD can be divided into the following points: As more member VPMCDs perform poorly on a feature set, the feature set would be more difficult. As more member VPMCDs perform well on a feature subset, the feature subset would be easier. In other words, the WV-VPMCD fused the knowledge acquired by local VMPCDs to make a consensus decision which is supposed to be superior to one attained by any individual VMPCD solely. In this method, the discrimination rate of member VPMCD on each feature subset and the irregular distribution of features in each subset are considered in the linear decision fusion of VPMCDs, so the classification output can be more reasonably obtained.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2943507, IEEE Transactions on Industrial Informatics

VI. EXPERIMENTAL RESULTS

In this section, the proposed method is evaluated on the public bearing datasets from Case Western Reserve Lab. The test bearing is SKF deep groove ball bearings with motor load about 3Hp and speed of 1797 rpm. Vibration signals were collected under four different operating conditions including normal, ball fault, inner race fault, and outer race fault. In each fault bearing, single point fault was set into the test bearing using electro-discharge machining with different defect sizes in diameter, i.e., 0.007, 0.14 and 0.021 inches, respectively. Thus, there were 10 conditions in total which are numbered in Table 2. An accelerometer with a bandwidth of up to 5000Hz was mounted on the motor housing at the drive end of the motor to collect the vibration signals from the bearing with motor speed of 1797 rpm. The sampling frequency was 12 kHz. A detailed signal collection depiction can be easily found in [26]. Because the data were stored as a long array, the data are divided into many short-time samples with 1024 data points in length.

According to Step 2 in Section 5, the values of m and τ, are calculated for the signals in different states and the corresponding RPs are shown in Fig.2. In this figure, the number and distribution of the recursion points (black points) basically vary with the fault state of rolling bearing. Any slow amplitude variation results in white area (the regions in which recurrences do not occur) visible on the RPs and while white circular regions may result from frequency shift inside the signal. In general, they are typical for non-stationary data and indicate that some stages are rare and transitions may occur. In Fig.2 (a)-(d), the transition from different fault types is associated with different vertical line structures. The apparent evolution from one state to another state like checker board and larger square patterns into a collection of vertical and horizontal short lines. The RPs for the same fault type with different defect sizes are provided in Fig. 2(d)-(f), where the density of recursive point decreases with the increase of defect size. For the big size of defect, the corresponding RP presents a massive aggregation pattern of Fig.2 (f). It is difficult to provide precise interpretation of the physical meaning of the white or black regions visible on the RPs. To quantify the state information in the RPs, the RQA features in Table 1 are extracted for all the fault states. Statistically speaking, the boxplots of these features are shown in Fig.3, in order to show the distribution of 11 features.

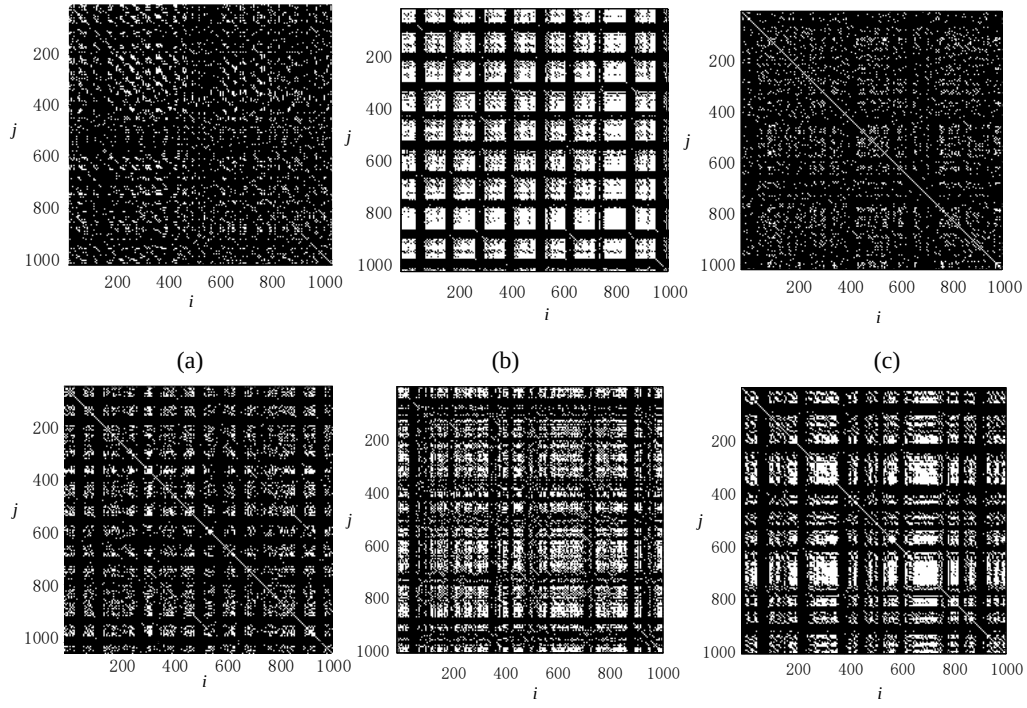
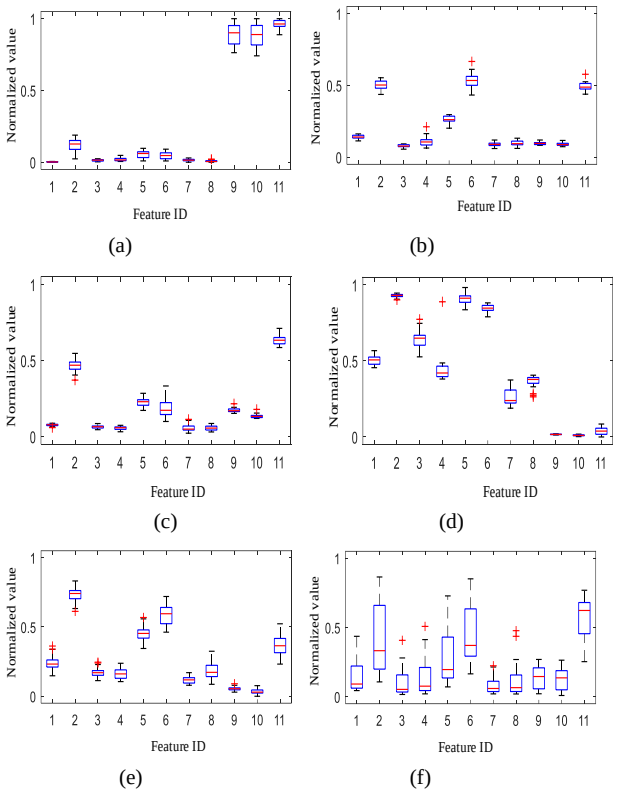


Fig.2 RPs of bearing faults. (a)Normal,(b)OR007, (c)B007, (d)IR007,(e)IR014,(f)IR021

In Fig.3 (a) and Fig.3 (g), the features have high similarity and their independence is not satisfactory. There exist more outlier samples in Fig. 3(f) and Fig.3 (i). The class aggregation of features in Fig.3 (f) and Fig.3 (j) is not good due to their discrete distributions. In addition, the overall distribution of RQA features in Fig.3 (g) is highly similar to that in Fig.3 (i). We can see that the RQA features have different discrimination ability, inconsistent stability and irregular distributions. In certain ranges, the more regular feature values it has, the more practicality there is. The feature similarity and irregular distribution mean more redundancy and less relevance of feature set, thereby resulting in the decrease in the stability and accuracy of VPMCD. Therefore, it is very necessary to regroup the features for improving the performance of VPMCD.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2943507, IEEE Transactions on Industrial Informatics



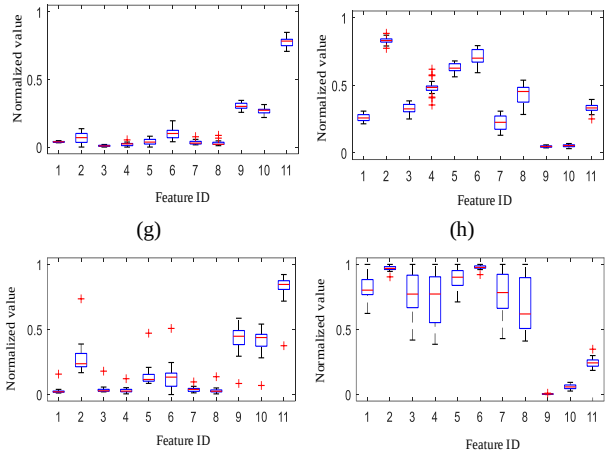


Fig. 3 Boxplots of RQA features, (a) Normal, (b)IR007, (c) B007,(d)OR007, (e) IR014, (f)B014,(g)OR014,(h)IR021, (i)B021, (j)OR021

Table II shows the VPM fitting errors and classifying results for 10 samples, when all eleven RQA features are put into the conventional VPMCD. This table shows that the sample data with "g" state label is predicted to be "i" state label. The misclassification phenomenon is very common in the cross validation. With the proposed method in this paper, all the features in Table I are ranked according to their RS, and the results are given in Table III.

TABLE II
RESULTS OF FAULT IDENTIFICATION BASED ON RQA AND VOTE-VPMCD

Fault type	Real state label	Square sum of prediction errors										Identification result
		VPM^1	VPM^2	VPM^3	VPM^4	VPM^5	VPM^6	VPM^7	VPM^8	VPM^9	VPM^{10}	
Normal	a	**7.1×10^{-5}**	13.91	30.97	1.8×10^{3}	1.331	0.038	36.54	7.951	12.31	433.4	a
IR 007	b	17.36	**6.7×10^{-5}**	1.530	404.4	0.134	9.1×10^{-3}	539.2	2.420	1.2×10^{4}	214.7	b
IR014	c	23.91	0.155	**5.6×10^{-3}**	783.2	0.439	0.181	239.9	5.139	659.9	7.7×10^{4}	c
IR021	d	7.4×10^{6}	1.6×10^{8}	6.2×10^{6}	**2.2×10^{-4}**	8.822	360.2	2.1×10^{8}	2.240	6.3×10^{8}	1.575	d
B007	e	0.572	0.043	16.42	1.1×10^{3}	**1.5×10^{-5}**	0.183	757.6	37.82	153.5	69.10	e
B014	f	1.476	0.152	0.605	372.0	0.134	**1.5×10^{-3}**	1.1×10^{3}	7.395	2.7×10^{3}	72.17	f
B021	g	0.194	1.8×10^{3}	251.9	9.3×10^{3}	1.284	6.8×10^{-3}	2.1×10^{-3}	528.9	**1.1×10^{-3}**	1.6×10^{3}	**i**
OR007	h	6.2×10^{4}	9.6×10^{5}	9.8×10^{3}	1.001	0.413	5.221	4.7×10^{6}	**1.4×10^{-3}**	1.3×10^{7}	9.136	h
OR014	i	0.269	475.5	74.77	8.0×10^{3}	0.325	0.114	0.035	367.5	**9.6×10^{-4}**	634.4	i
OR021	j	1.7×10^{8}	4.4×10^{9}	2.7×10^{8}	60.93	3.3×10^{3}	1.2×10^{5}	3.6×10^{9}	481.2	1.0×10^{10}	**4.6×10^{-3}**	j

Note: In the fault type, IR, B and OR represent inner race fault, rolling ball fault and outer race fault, respectively. 007, 014, 021 represent the defect size 0.007inch, 0.014inch, 0.021inch, respectively.

TABLE III
RESULTS OF RSAP

Feature ID	Name	RS	NRS	Cluster ID	Subset ID
1	RR	0.231	0.086	FC_1	$\mathbf{FS}_2$
2	DET	0.323	0.120	FC_2	$\mathbf{FS}_1$
3	L	0.237	0.088	FC_2	$\mathbf{FS}_1$
4	L_{max}	0.195	0.0727	FC_2	$\mathbf{FS}_3$
5	ENTR	0.315	0.117	FC_1	$\mathbf{FS}_3$
6	LAM	0.318	0.119	FC_1	$\mathbf{FS}_2$
7	TT	0.160	0.059	FC_2	—
8	V_{max}	0.167	0.062	FC_2	—
9	T_j^1	0.236	0.087	FC_3	$\mathbf{FS}_2$
10	T_j^2	0.221	0.082	FC_3	$\mathbf{FS}_3$
11	RET	0.280	0.104	FC_3	$\mathbf{FS}_1$

In Table III, each feature has its own RS that presents its priority for discriminating fault states. According to the results of AP clustering, the features are automatically divided into three clusters (FC_1, FC_2 and FC_3). The features in the same cluster generally have maximum similarity and big relevance, so it is better to separate them to avoid the information redundancy. Using the RSAP method, the NRS top features are selected from each cluster to form a new feature subset, i.e. feature 2, 3 and 11 form $\mathbf{FS}_1$. The features ranking second in each cluster forms feature subset 2, i.e. feature 1, 6 and 9 form $\mathbf{FS}_2$. In the same way, feature 4, 5, 10 form $\mathbf{FS}_3$. Owing that the NRS of feature 7 and 8 are very small, they are abandoned in the VPM establishment. Each subset is put into a resampling procedure to generate a set of training resampled dataset. After that, each resampled dataset corresponds to a sub-VPM. For

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2943507, IEEE Transactions on Industrial Informatics

each sub-VPM, the un-sampled instances are used as the input to obtain the corresponding discrimination rate. The classifiers are weighted according to their discrimination rates. All of the subset $\mathbf{FS}_1$, $\mathbf{FS}_2$ and $\mathbf{FS}_3$ are processed with bagged ensemble VPM algorithm and their outputs are fused into the final identification result.

To further evaluate the performances of proposed method, the same training and test samples are used to carry out the conventional VPMCD, SVM and BP neural network identification test. The comparisons of identification accuracy and computational cost are described as Fig.4 (a) and (b). The average identification accuracies of WV-VPMCD, SVM, VPMCD and BP are 96.96%, 91.79%, 90.19%, 87.89%, respectively. Their corresponding average computational time consumption are 3.225s, 16.225s, 0.147s and 0.795s, respectively. From Fig.4 (a), the identification accuracies of all the classifiers increase along with the number of training samples. The accuracy curve of conventional VPMCD is very steep, which means that the number of training samples seriously affects its identification results. The WV-VPMCD yields higher accuracy and outperforms the other algorithms. Moreover, the curve of WV-VPMCD tends to be gentle, which indicates its relative independency to the number of training sample. In Fig.4 (b), the computational cost of WV-VPMCD is higher than that of the conventional VPMCD and BP, but less than that of SVM. Owing to the multiple regression in VPMCD with different Bagging sets, the accuracy of WV-VPMCD increases at the expense of the computation cost to some extent, but the cost is acceptable.

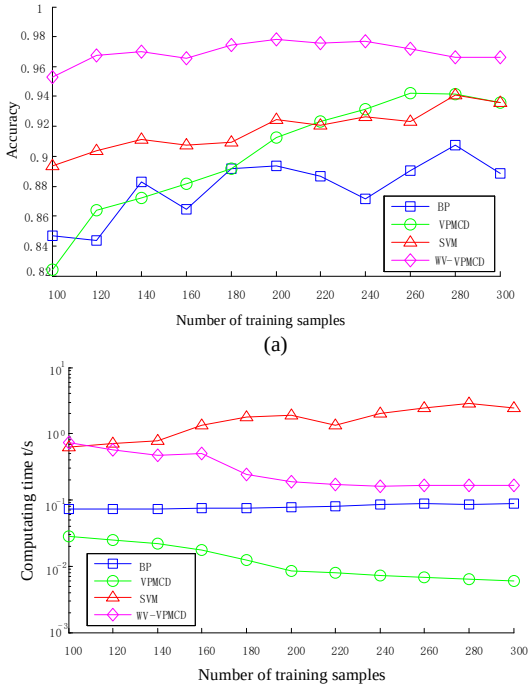


Fig.4 Comparisons of BP, VPMCD, SVM and WV-VPMCD . (a) Identification accuracy, (b) Time consumption

TABLE IV
PREDICTION RESULTS OF VPMCD IN SMALL SAMPLES

Real state	Identification results											
	1	2	3	4	5	6	7	8	9	10	Majority voting	discrimination rate
a	a	a	**f**	a	**b**	a	a	a	**g**	a	a	7/10
b	b	b	b	**a**	b	b	b	b	b	**f**	b	8/10
c	**g**	c	c	c	c	c	**e**	**i**	c	c	c	7/10
d	d	**j**	d	d	d	**j**	d	d	**h**	d	d	7/10
e	**f**	e	e	**f**	e	**a**	e	**c**	**f**	**f**	unable	4/10
f	f	**b**	**a**	f	**b**	f	f	**e**	f	**c**	f	5/10
g	f	**i**	**i**	g	**i**	g	**i**	g	**i**	g	**i**	**4/10**
h	**d**	h	h	h	**d**	**f**	h	h	h	h	h	7/10
i	i	**g**	i	i	**c**	i	**G**	**g**	i	i	i	6/10
j	j	j	j	**d**	j	**d**	j	j	j	j	j	8/10

In order to verify the identification performance of WV-VPMCD in the case of small training samples. 4 to 10 samples are extracted from each state feature subset for training. For ten kinds of bearing fault states namely, 40 to100 training samples are used to calculate the discrimination rate. In the case of 100 training samples (10 samples for each state), the corresponding identification results are provided in Table IV, using the conventional VPMCD without feature selection. In this table, there is a lot of uncertainty in the identification results if with one single VPMCD. And even with the method of majority voting, there still exists the phenomenon of misclassification. That is because the VPM parameters cannot be precisely estimated out by least square fitting when the number of training samples is not much bigger that of model parameters. And some important model parameters would be abandoned, which leads poor fitting results, especially for higher-order interaction model.

Fig.5 provides the accuracy comparison of different classifiers in the case of small training samples. If without enough sample features, the BP and SVM are easy to relapse into the local extremum. Both of them cannot effectively differentiate the feature sets with overlapping patterns and their identification accuracies keep about 80%, as shown in Fig.5. More serious misclassification phenomenon exists in the conventional VPMCD and its accuracy is always below 80%. As analyzed in Section III, in the case of small training samples only weak VPM can be obtained, which represents part of the feature interaction relationship. In the WV-VPMCD, the partial interaction relationships are integrated by assigning the optimal weights according to the accuracy of individual VPMCD, and then are fully utilized to identify fault states. Furthermore, the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2943507, IEEE Transactions on Industrial Informatics

weighted bagging ensemble algorithm divides the irregularly distributed data into the regular feature subsets for training sub-VPMs so as to improve the stability of WV-VPMCD. Therefore, the WV-VPMCD always keeps a good accuracy above 90% even with 40 feature samples. The results demonstrate that the proposed method achieves high classification accuracy and good stability for intelligent fault diagnosis.

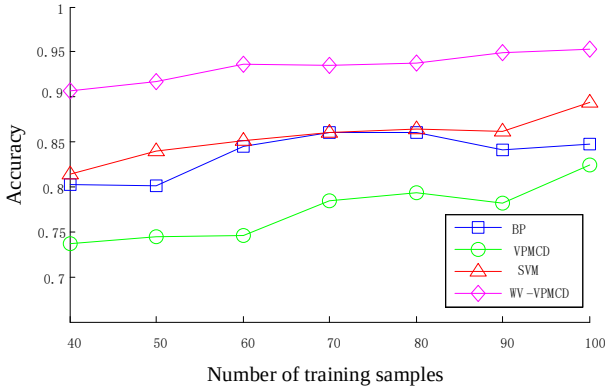


Fig.5 Identification accuracy of BP, VPMCD, SVM and WV-VPMCD with small samples

VII. Conclusion

The study presented a novel intelligent fault diagnosis method based on the proposed WV-VPMCD. This method does not involve the optimization and iteration of parameters, and thus decreases the computational cost of intelligent learning. In this method, RQA features are extracted to characterize the nonlinear short time series and divided into the different subsets of non-redundant features with the same level of discrimination ability. Each feature subset is used to train individual sub-VPM. The weight of each sub-VPM is assigned according to its individual discrimination rate, thus resulting in a flexible exploitation of complementary competencies of sub-VPMs. The WV-VPMCD method integrates the mutual interaction relationships of all the features and considers the uncertainty of selecting training samples. It does not need to rely on prior knowledge and regular samples, therefore it is less influenced by subjective factors and singular samples perturbation to get the objective and stable classification results, even in the case of small samples and high-dimension feature vector. The experimental results have validated the effectiveness and superiority of the method for bearing fault diagnosis. Efforts will also focus on verifying the method's robustness to some other external disturbances, such as rotating speed, operating load and output power, etc.

References

[1]. M Cococcioni, B Lazzerini, and SL Volpi, "Robust Diagnosis of Rolling Element Bearings Based on Classification Techniques," IEEE Transactions on Industrial Informatics, Vol.9, no.4, Nov.2013, pp. 2256-2263.

[2]. R Singleton， E Strangas， S Aviyente, "The Use of Bearing Currents and Vibrations in Lifetime Estimation of Bearings," IEEE Transactions on Industrial Informatics, Vol.13, no.3, pp.1301-1309, 2017

[3]. M Van, HJ Kang, "Bearing defect classification based on individual wavelet local fisher discriminant analysis with particle swarm optimization," IEEE transactions on industrial informatics, Vol.21, No.1, pp. 124-135, 2016.

[4]. Z Gao, C Cecati and SX Ding, "A survey of fault diagnosis and fault tolerant techniques—part II: Fault diagnosis with knowledge-based and hybrid/active approaches," IEEE Transactions on Industrial Electronics, Vol.62, no.6, pp. 3768–3774 .2015

[5]. CA Perez-Ramirez, JP Amezquita-Sanchez and M Valtierra-Rodriguez, et al. "Fractal dimension theory-based approach for bearing fault detection in induction motors." IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC) 2017:1-6, Ixtapa, MEXICO, Nov. 09-11, 2017.

[6]. L Fu, Y Wei and S Fang et al. "Condition Monitoring for Roller Bearings of Wind Turbines Based on Health Evaluation under Variable Operating States," Energies, Vol.10, no.10, pp.1564-1675, 2017

[7]. L Han, C Li and H Liu. "Feature Extraction Method of Rolling Bearing Fault Signal Based on EEMD and Cloud Model Characteristic Entropy," Entropy, Vol.17, no.10, pp. 6683-6697, 2015.

[8]. K Zhu, T Mei, DS Ye. "Online condition monitoring in micro-milling: A force waveform shape analysis approach", IEEE Transactions on Industrial Electronics: 2015, Vol.62, no.6, pp.3806-3813, 2015.

[9]. N Marwan. "A historical review of recurrence plots," The European Physical Journal Special Topics, Vol. 164, no.1, pp.3-12, 2018.

[10]. G Litak and R Longwic. "Analysis of repeatability of Diesel engine acceleration," Applied Thermal Engineering, Vol.29, no.17, pp. 3574-3578, 2010.

[11]. G Litak, M Borowiec, and J Hunicz etc. "Vibrations of a delivery car excited by railway track crossing," Chaos Solitons & Fractals, Vol. 42, no.1, pp.270-276, 2009.

[12]. X Liu and L Bo. "Identification of Resonance States of rotor-bearing system using RQA and Optimal Binary Tree SVM", Neurocomputing, Vol.152, no.152, pp. 36-44, 2015.

[13]. M Elforjani, S Shanbr. "Prognosis of bearing acoustic emission signals using supervised machine learning," IEEE Transactions on Industrial Electronics, Vol.65, No.7, pp.5864-5871, 2018.

[14]. L Wen, X Li, L Gao, Y Zhang. "A new convolutional neural network-Based Data-Driven fault diagnosis Method," IEEE Trans. Ind. Electron., Vol. 65, no.7, pp: 5990-5998, JUL 2018

[15]. K Zhu, T Liu, On-line High-Speed Milling Tool Wear Monitoring via Hidden Semi-Markov Model with Dependent Durations, IEEE Transactions on Industrial Informatics, Vol,14, no.1, pp. 69–78, 2018.

[16]. R Raghuraj, S. Lakshminarayanan. "Variable predictive model based classification algorithm for effective separation of protein structural classes," Computational Biology and Chemistry, Computational Biology and Chemistry, Vol. 32, no.4, pp. 302–306, 2008.

[17]. Y. Yang, H.Y. Pan, L. Ma and J.S. Cheng, "A roller bearing fault diagnosis method based on the improved ITD and RRVPMCD," Measurement, Vol.55 pp.255-264, 2014

[18]. SR Luo, JS Cheng, KX Wei, "A Fault Diagnosis Model Based on LCD-SVD-ANN-MIV and VPMCD for Rotating Machinery", Shock and Vibration, (2016-9-20), pp. 1-10, 2016.

[19]. T Tao, Lin B, XF Liu, B Sun, DP Wei, "Variable predictive model class discrimination using novel predictive models and adaptive feature selection for bearing fault identification," Journal of Sound and Vibration. Vol. 425, pp.137-148, 2018.

[20]. SR Luo, JS Cheng, M Zeng and Y Yang, "An intelligent fault diagnosis model for rotating machinery based on multi-scale higher order singular spectrum analysis and GA-VPMCD," Measurement, Vol.87, pp38-50, 2016.

[21]. YF Jia, YL Zhu, "Partial discharge pattern recognition using variable predictive model-based class discrimination with kernel partial least squares regression," IET Science, Measurement & Technology, Vol.12, pp.360-367, 2018.

[22]. E Hancer, B Xue, M Zhang. "Differential evolution for filter feature selection based on information theory and feature ranking" Knowledge-Based Systems, Vol.140, no.5, pp.103-119, 2018

[23]. U Bodenhofer, A Kothmeier, S Hochreiter. "AP Cluster: an R package for affinity propagation clustering," Bioinformatics, Vol. 27, no.17, pp. 2463±4, 2011.

[24]. N. Marwan. "How to avoid potential pitfalls in recurrence plot based data analysis". International Journal of Bifurcation and Chaos, Vol.21, no.4,pp.1003-1017, 2011

[25]. HDI Abarbanel, N Masuda, MI Rabinovich, ETumer, "Distribution of Mutual Information," Phys Lett A, Vol. 281, pp.368-373, 2012.

[26]. http://csegroups.case.edu/bearingdatacenter/pages/download-data-file