Journal of Sound and Vibration 425 (2018) 137–148

Contents lists available at ScienceDirect

Journal of Sound and Vibration

journal homepage: www.elsevier.com/locate/jsvi

Variable predictive model class discrimination using novel predictive models and adaptive feature selection for bearing fault identification

Tao Tang, Lin Bo*, Xiaofeng Liu, Bing Sun, Daiping Wei

The State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing 400044, PR China

ARTICLE INFO

Article history:
Received 3 November 2017
Received in revised form 1 March 2018
Accepted 30 March 2018
Available online 9 April 2018
Handling Editor: K. Shin

Keywords:
Variable predictive model
Rolling bearing diagnosis
Neural-network
Feature selection
Affinity propagation
RReliefF

ABSTRACT

A complete fault diagnosis for the rolling bearing is proposed in this paper. Variable predictive model class discrimination (VPMCD) is a conventional pattern recognition method; however, in practice, when the fault diagnosis method is applied to small samples or in multi-correlative feature space, the stability of the VPM constructed based on the least squares (LS) method is not sufficient. Based on affinity propagation (AP) clustering, RReliefF, and sequential forward search, the ARSFS is proposed to select the significant subset of original feature set and to reduce the dimension and multiple correlations of the feature space. Further, this paper uses two kinds of Gaussian Neural Network, namely the Radial Basis Function Neural Network (RBF) and the Generalized Regression Neural Network (GRNN), instead of the LS method to construct predictive models of VPMCD, called AOR-VPMCD. Compared with the conventional VPMCD and its improvements, based on sufficient experiments, the entire process presented in this paper can effectively identify the fault of the rolling bearing.

1. Introduction

Vibration signals gathered from bearings usually involve a great deal of information about the state of the machine's operation [1,2]. The variable predictive model-based class discrimination as a distinct method of pattern recognition was proposed by Raghuraj, R et al. [3], and this method has been wildly used in fault diagnosis of the rolling bearing. According to the significant differences of the intrinsic correlation between characteristics in different bearing states, different samples can be distinguished. Compared with other algorithms, VPMCD avoids the iterative process of the artificial neural network (ANN) [4] and the optimization process in the SVM [5], and there are no preset parameters. However, practically in the face of the varied effects of non-linear factors such as load, friction, clearance, and stiffness, it is difficult to establish precise variable predictive models (VPMs) with merely linear or quadratic models.

Aiming at the shortcomings of the conventional models in abnormal data processing, a robust regression-variable predictive model are proposed by Yu Yang et al. [6], which makes up the problem that least square method is sensitive to abnormal data. Similarly, Songrong Luo in Ref. [7] combined ANN with the mean impact value (MIV) to reduce feature dimensions and enhance the stability of the VPMCD output. Nevertheless, they both ignore VPMCD estimated by the Least

* Corresponding author.
E-mail address: bolin0001@aliyun.com (L. Bo).

https://doi.org/10.1016/j.jsv.2018.03.032
0022-460X/

Squares (LS) method [8] is sensitive not only to the abnormal data but also to the existing multi-correlation between variables. Recently by Cui et al. [9], Partial Least Squares method (PLS) instead of the LS method was used to estimate the model parameters of VPMCD to address the negative influence of multiple correlations. Besides, to strengthen the robustness of recognition, GA-VPMCD was put forward by combining the genetic algorithm (GA) with four conventional VPMs by Luo SR et al. [10]. It does break the limitations of one single VPM, but the GA method can also contribute to considerable time costs.

To overcome these deficiencies, an adaptive optimal radial model-based VPMCD is proposed, called AOR-VPMCD. The new VPMs are constructed by two artificial neural networks, namely the Radial Basis Function Neural Network (RBF) [11] and the Generalized Regression Neural Network (GRNN) [12]. Unlike other methods of improvement, we managed to redefine the new VPMs rather than strengthening the original ones and to verify its superiority by comparing other improvements.

These features extracted from the original or preprocessing signals may have multiple correlations or large dimensions, which is not conducive to pattern recognition in VPMCD. Most of the feature selection (FS) methods can be roughly divided into two broad categories, namely, filter and wrapper. Regarding filter methods, features are scored and graded according to statistical criteria. For example, in Ref. [13], Zexian Wei et al. have designed a feature selector based on affinity propagation (AP) clustering algorithm to extract the cluster centers as optimal features at the cost of considerable information loss, as just these central points may be insufficient to carry all the state information. In Ref [14], considering the varied contributions of features, Abrahamsson, V et al. combined F-statistics feature selection with ReliefF algorithm, but the artificial thresholds have great randomness. On the wrapper methods, Alrajab, M et al. [15] proposed a three-phase combinatorial feature selection algorithm, including particle swarm optimization (PSO) and Support Vector Machine (SVM) algorithm. Although this method has been identified to do well in feature selection, it also has higher computational costs.

This paper uses sequential forward selection (SFS) method and two filtering feature selection methods, namely AP [16] and RReliefF [17], to create our unsupervised feature selector called ARSFS. First, the raw features are evaluated by the RReliefF algorithm and are clustered into several subsets using AP clustering, which is to distinguish the useful correlation from multi-correlated features. Second, based on AP clustering and SFS method, a feature evaluation loop is designed to extract key features from each subset. Features extracted from all feature subsets can inherit the correlation between subsets. Without any manual parameters, ARSFS can use feature selection loops to differentiate between critical and unrelated features. Besides, the number of key features can be determined automatically. For further comparison, several feature methods have been applied to verify the ability of ARSFS, such as the feature center method and ReliefF-SFS method.

The remainder of this paper is organized as follows. In Section 2, the fundamental theories are briefly introduced, including feature extraction, AP, and RReliefF. In Section 3, the proposed feature selector is illustrated. Section 4 summarizes the VPMCD and AOR-VPMCD method respectively. Section 5 presents and analyses the engineering application of the proposed techniques. Finally, some conclusions are brought in Section 6.

2. Brief review of basic theories

2.1. Feature extraction

Recurrence Quantification Analysis (RQA) [18,19] is a method of nonlinear time series analysis. The primary focus lies on recurrence plots (RP) and their quantification. Wherein RPs are briefly defined as

$$\mathbf{R}_{i,j}^{m,\varepsilon_i} = H\left(\varepsilon_i - \left\|\vec{\mathbf{x}}_i - \vec{\mathbf{x}}_j\right\|\right), i,j \in [1, N-(m-1)\tau] \tag{1}$$

where $H()$ is the Heaviside function, and ε is a predefined threshold (1.3 times the standard deviation of the series), and $\vec{\mathbf{x}}_\mathbf{i}$, $\vec{\mathbf{x}}_\mathbf{j}$ are phase space trajectories in m-dimension phase space [20,21], these trajectories can be reconstructed from single time series by using a time delay τ [22,23].

In this study, 11 RQA parameters (listed in Table 1) are used to characterize those non-linear and non-stationary bearing states, and detailed parameters are listed in Table 3.

Table 1
The extract eleven-feature parameters.

Feature parameters	Equation	Feature parameters	equation
1.Recurrence Rate	$RR = \frac{1}{N^2}\sum_{i,j=1}^{N}\mathbf{R}_{i,j}^{m,\varepsilon}$	7.Trapping Time	$TT = \sum_{v=v_{\min}}^{N} v\mathbf{P}^{e}(v)/\sum_{v=v_{\min}}^{N}\mathbf{P}^{e}(v)$
2.Determinism	$DET = \sum_{l=l_{\min}}^{N} l\mathbf{P}^{e}(l)/\sum_{i,j=1}^{N}\mathbf{R}_{i,j}^{m,\varepsilon}$	8.Maximal vertical line length	$V_{\max} = \max(\{v_i, i = 1,2,\cdots,N_v\})$
3.Mean diagonal line length	$L = \sum_{l=l_{\min}}^{N} l\mathbf{p}(l)/\sum_{l=l_{\min}}^{N}\mathbf{p}(l)$	9.Recurrence Time of 1st type	$T_i^1 = \left\lvert\{i,j : \vec{\mathbf{x}}_i, \vec{\mathbf{x}}_j \in \Re_i\}\right\rvert$
4.Maximal diagonal line length	$L_{\max} = \max(\{l_i, i = 1,2,\cdots,N_l\})$	10.Recurrence Time of 2nd type	$T_j^2 = \left\lvert\{i,j : \vec{\mathbf{x}}_i, \vec{\mathbf{x}}_j \in \Re_i; \vec{\mathbf{x}}_{i-1} \notin \Re_i\}\right\rvert$
5.Entropy	$ENTR = -\sum_{l=l_{\min}}^{N} v\mathbf{p}(l)\ln\mathbf{p}(l)$	11.Recurrence Time Entropy	$RTE = -\mathbf{p}(TT)\ln\mathbf{p}(TT)$
6.Laminarity	$LAM = \sum_{v=v_{\min}}^{N} v\mathbf{P}^{e}(v)/\sum_{v=1}^{N} v\mathbf{P}^{e}(v)$		

Notes: 1. $\mathbf{P}^{e}(l) = \{l_i, i = 1,2,\cdots,N_l\}$ is the frequency distribution of the lengths l of diagonal structures; 2. $\mathbf{P}^{e}(v) = \{v_i, i = 1,2,\cdots,N_v\}$ is the frequency distribution of the lengths l of vertical structures; 3. N_l and N_v are the absolute number of diagonal and vertical lines, respectively; 4. $\mathbf{p}(l) = \mathbf{P}^{e}(l)/\sum_{l=l_{\min}}^{N}\mathbf{P}^{e}(l)$; 5. $\Re_i$ denotes the recurrence points which belong to the trajectory $\vec{\mathbf{x}}_i$.

2.2. Affinity propagation cluster

The Affinity propagation method (AP) was proposed by Frey et al. [16], which simultaneously treats all data points as possible centers in the network, and real-value messages are transmitted along the edges of the network until a set of suitable centers and corresponding clusters are found. There are two types of messages in the AP algorithm, namely availability message ($a(i,k)$) and responsibility message ($r(i,k)$).

The responsibility message, sent from data point i to the candidate exemplar point k, indicates how appropriate for the point i determines point k as its exemplary. Responsibility information update rules are as follows:

$$r(i,k) \leftarrow s(i,k) - \max_{k' \neq k}\{a(i,k') + s(i,k')\} \tag{2}$$

$$a(i,k) = \begin{cases} \min\left\{0, r(k,k) + \sum\limits_{i' \text{s.t.} i' \in \{i,k\}} \max\{0, r(i',k)\}\right\}, i \neq k \\ \sum\limits_{i' \text{s.t.} i' \in \{k\}} \max\{0, r(i',k)\}, i = k \end{cases} \tag{3}$$

where the $a(i,k)$ denotes how properly the point k is identified as the exemplar of point k. After a number of iterations or iterative convergence is reached, each point would select its center by $a(i,k) + r(i,k)$ value. Detailed insights into the AP algorithm can be found in Refs. [24,25]. This procedure is applied not only to identify the features but also to identify the fault categories in Section 3.

3. Adaptive feature selection (ARSFS)

Based on AP clustering, RReliefF and SFS method, the detail of feature selection method (ARSFS) is illustrated in Fig. 1. For the sake of illustration, 200 samples (20 samples per category) are taken into consideration. $\mathbf{S}_{11\times 200}$ is the initial feature space that contains 11 features and 200 samples per feature.

In the 1st step, RReliefF method is used to calculate the contribute of each feature to pattern recognition. RReliefF as a classical feature evaluation algorithm for regression problems was introduced in Ref. [26]. It estimates the discriminative power of each feature F_j between adjacent instances by approximating the following difference of probabilities:

$$RF_j = \frac{P_{(diff\ C|diff\ F_j)} P_{diff\ F_j}}{P_{diff\ C}} - \frac{\left(1 - P_{(diff\ C|diff\ F_j)}\right) P_{diff\ F_j}}{1 - P_{diff\ C}} \tag{4}$$

where $P_{diff\ C}$ and $P_{diff\ F_j}$ indicate the probability that two nearest instances have different predictions, or have different values for F_j, respectively. And $P_{(diff\ C|diff\ F_j)}$ denotes the probability that two nearest cases have different predictions and different values for F_j.

Investigating the feature selection flow in Fig. 1, we can see that 11 raw features are clustered into three subsets and listed in descending order by RFs in the 1st step. These feature subsets can help to distinguish homogenous features, and in each subset, the order of each feature shows the degree of association with the bearing condition category.

In the 2nd step, based on the traversal algorithm: SFS algorithm [27,28], the pattern recognition accuracy of AP classification is used as the evaluation criterion, and the key features are selected from each subset:

$$\tilde{\mathbf{S}}_d = \tilde{\mathbf{S}}_{d-1} \cup \arg\max_{F_i} \mathrm{SFS}\left(\tilde{\mathbf{S}}_{d-1} \cup F_i, L, M\right) \tag{5}$$

where $\tilde{\mathbf{S}}_d$ refers to the optimal feature set containing d features. L indicates the class labels for each bearing states. M is the classification model (AP clustering).

As shown in Fig. 5, the entire process starts with the initial subset $\tilde{\mathbf{S}}$ composed of the first feature of each subset. The entire traversal process begins with the subset whose second feature can maximise recognition accuracy. For example, in this article, when adding the second feature of the subset 3 (F9) into $\tilde{\mathbf{S}}$, the accuracy increased by 19.5%, which is better than that of F_1 (8.5%) and F_6 (12%), so we start with Subset 3. The loop then iteratively updates $\tilde{\mathbf{S}}$ by adding feature variable F_i until no significant improvement in classification performance is observed ($<0.5\%$) [29] and jumps into next subset.

4. VPMCD-based novel predictive model

4.1. VPMCD method introduction

It is known that interactions among feature variables have different performance between different bearing states. In VPMCD, a set of VPMs such as Linear (L), Linear and Interaction (LI), Quadratic (Q), Quadratic and Interaction (QI) model had been provided by Raghuraj, R. [3], in the form of Eqs. (6)–(9).

$$\mathrm{L-VPM}: F_i = b_0 + \sum_{j=1}^{r} b_j F_j \tag{6}$$

$$\mathrm{LI-VPM}: F_i = b_0 + \sum_{j=1}^{r} b_j F_j + \sum_{j=1}^{r} \sum_{k=j+1}^{r} b_{jk} F_j F_k \tag{7}$$

$$\mathrm{Q-VPM}: F_i = b_0 + \sum_{j=1}^{r} b_j F_j + \sum_{j=1}^{r} b_{jj} F_j^2 \tag{8}$$

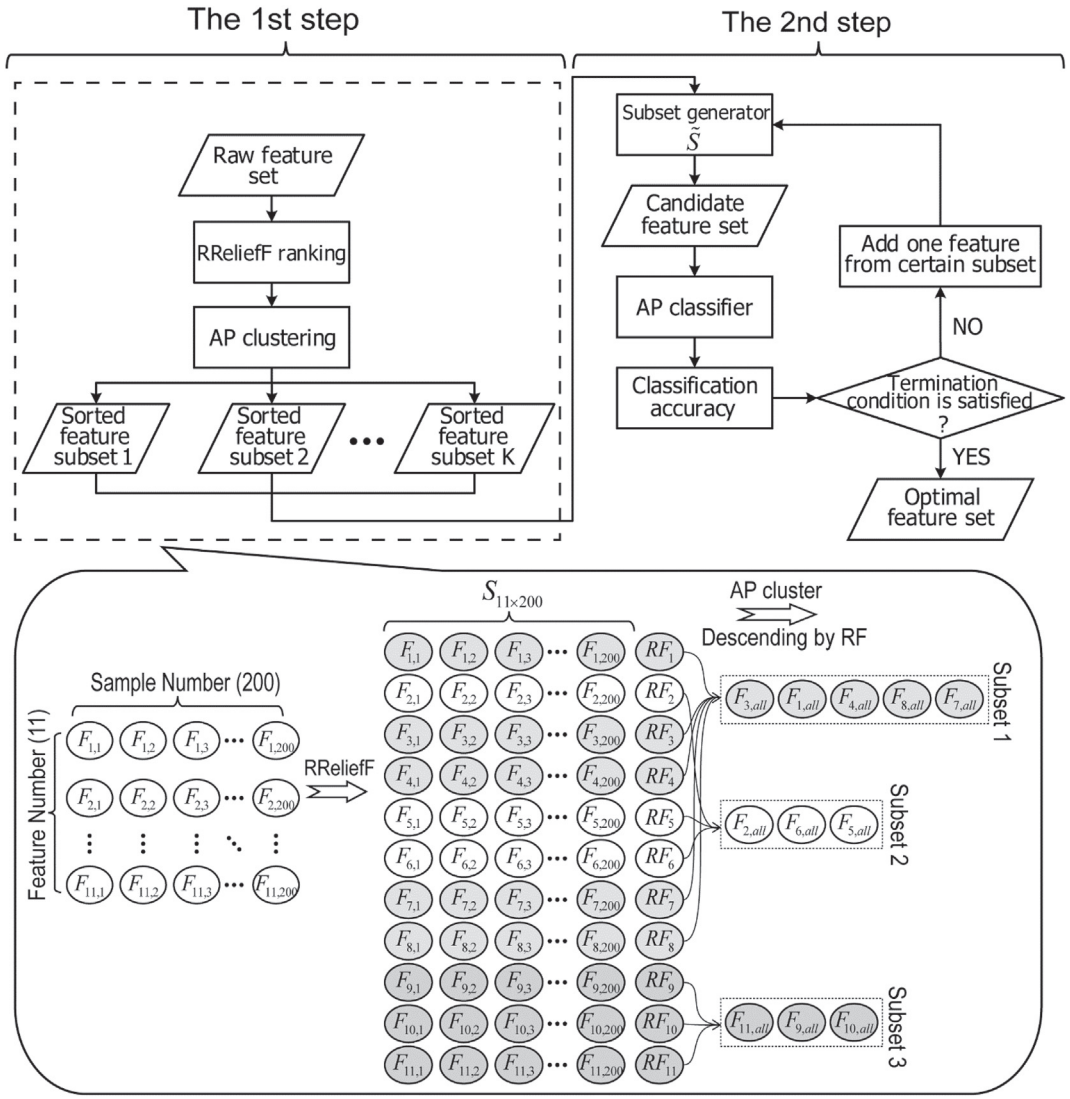

Fig. 1. Flowchart of the ARSFS method.

$$\mathrm{QI-VPM}: F_i = b_0 + \sum_{j=1}^{r} b_j F_j + \sum_{j=1}^{r} b_{jj} F_j^2 + \sum_{j=1}^{r} \sum_{k=j+1}^{r} b_{jk} F_j F_k \tag{9}$$

Suppose there are g classes and m different feature variables per class which can be expressed by a vector $\mathbf{S} = \{F_1, F_2, \cdots, F_m\}$. After selecting one of the above regression models, VPM_i for any feature variables F_i can be modeled via using measurements of the remaining variables $F_j (j = 1, 2, \cdots, m, j \neq i)$. It is noted that the model with minimum predictive errors during validation is selected as optimal VPM_i for feature variable F_i, as well as the collection of these optimal $\mathrm{VPM}_i (i = 1, 2, \cdots, m)$ is treated as an interaction model for representing the internal association.

4.2. AOR-VPMCD algorithm

Based on the structure of VPMCD, two new VPMs are defined in Table 2, namely RBF-VPM and GRNN-VPM. In this paper, they both are collectively called AOR-VPM owing to their similar effects and complementarities. More details about RBF and GRNN can refer to [30–32]. For any dependent variable F_i, it can be predicted by other independent variables $F_j (j = 1, 2, \cdots, m, j \neq i)$.

The steps of the classification recognition method based on AOR-VPMCD are as follows:

(1) Suppose there are training data covering g classes with each class comprising $n_k (k = 1, 2, \cdots, g)$ training data. Wherein $\mathbf{X} = \{X_1, X_2, \cdots, X_m\}$ is the feature vector comprising m features.
(2) For the k classification with m features, we can find corresponding model for each feature variable. The following equation called the k-th class predictive model set $\mathbf{AOR}_{m\times k}^{\mu}$.

$$\mathbf{AOR}_{m\times k}^{\mu} = \left[\mathbf{AOR}_{1,k}^{\mu}, \mathbf{AOR}_{2,k}^{\mu}, \cdots, \mathbf{AOR}_{m,k}^{\mu}\right]^T; \mu = 1, 2 \tag{10}$$

Where μ denotes the sequence number of the two models, namely $\mathrm{AOR}^{\mu=1}$ is the GRNN-VPM and $\mathrm{AOR}^{\mu=2}$ is the RBF-VPM.

(3) Let's start with $k = 1$ and compute squared prediction errors (SSE) of all models, and then select the model with the least SSE based on training vector set $\mathbf{X}$, where μ is determined to be a certain value.
(4) Let $k = k + 1$ and repeat step 3 until $k = g$, then stop.
(5) The model matrix $\mathbf{AOR}_{i,k} (i = 1, 2, \cdots, m; j = 1, 2, \cdots, g)$ which consists of separate predictive model for each feature has been found, where i stands for the different feature variable, and k is the class label.

$$\left[\mathbf{AOR}_{i,k}\right]_{i,k=1}^{m,g} = \begin{bmatrix} \mathrm{AOR}_{1,1} & \mathrm{AOR}_{1,2} & \cdots & \mathrm{AOR}_{1,g} \\ \mathrm{AOR}_{2,1} & \mathrm{AOR}_{2,2} & \cdots & \mathrm{AOR}_{2,g} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{AOR}_{m,1} & \mathrm{AOR}_{m,2} & \cdots & \mathrm{AOR}_{m,g} \end{bmatrix} \tag{11}$$

(6) For any test sample $\mathbf{X} = \{X_1, X_2, \cdots, X_m\}$, there will be g prediction values for each feature variable $X_i (i = 1, 2, \cdots, m)$ calculated by other features $X_j (j = 1, 2, \cdots, m; j \neq i)$ via Eq. (11) and (12).

Table 2
Novel variable predictive model.

	Equation	Parameters
GRNN-VPM	$\tilde{F}_i = \frac{\sum_{k=1}^{K} W_k \exp(-\|F_j - C_k\|^2 / 2\xi^2)}{\sum_{k=1}^{K} \exp(-\|F_j - C_k\|^2 / 2\xi^2)}$	a) K is the number of training patterns; b) $C_k$is the k-th neuron nod; c) W_k is the weighting factor; d) ξ is the diffusion factor;
RBF-VPM	$\tilde{F}_i = \sum_{n=1}^{N} \omega_n \exp\left(-\frac{\|F_j - C_n\|}{2\sigma_n^2}\right) + \omega_0$	a) N is center number; b) ω_n is the connection weight; c) ω_0 is the connection deviation; d) center C can be randomly selected from the training set;

Table 3
Description of datasets of bearing.

Class label	Fault type	Fault size (in.)	Number of training/testing samples	Abbreviation	RQA parameters	
					τ	m
1	Normal	0	60/30	Normal	1	5
2	Ball	0.007	60/30	B (0.007)	1	4
3	Ball	0.014	60/30	B (0.014)	1	4
4	Ball	0.021	60/30	B (0.021)	1	5
5	Inner race	0.007	60/30	IR (0.007)	1	4
6	Inner race	0.014	60/30	IR (0.014)	1	5
7	Inner race	0.021	60/30	IR (0.021)	1	4
8	Outer race	0.007	60/30	OR (0.007)	1	5
9	Outer race	0.014	60/30	OR (0.014)	1	4
10	Outer race	0.021	60/30	OR (0.021)	1	4

$$\tilde{\mathbf{X}}_{i,k} = \left[\mathrm{AOR}_{i,1}(X_j), \mathrm{AOR}_{i,2}(X_j), \cdots, \mathrm{AOR}_{i,g}(X_j)\right]; i = 1, 2, \cdots, m; k = 1, 2, \cdots, g \tag{12}$$

Where the $\tilde{\mathbf{X}}_{i,k}(k = 1, 2, \cdots, g)$ denotes g optimal estimates for the test sample, in which k represents the class label, and i represents the feature label.

(7) If the kth column predicted vector $\tilde{\mathbf{X}}_k$ could meet the following need (Eq. (13)) which has the least SSE^k. This sample would be classified to class k.

$$\min\left\|\mathrm{SSE}^k\right\| = \left\|\sum_{i=1}^{m}\left(X_i, \tilde{\mathbf{X}}_{i,k}\right)\right\|; k = 1, 2, \cdots, g \tag{13}$$

5. Experimental results

5.1. Overview of the experiments

The bearing vibration signals from Case Western Reserve University bearing Data Center are adopted to verify the effectiveness of proposed methods for rolling bearing fault identification. The test rig is shown in Fig. 2, mainly comprising a three-phase induction motor, accelerometer sensors, and a load motor. Each bearing is tested without load, and single point faults of sizes 0.007, 0.014 and 0.021 inches are set on the drive end bearings. These faults are set on the ball, inner and outer race, respectively. The motor speed is set at 1797 RPM, and the sampling rate is 12 kHz.

There are ten bearing condition patterns, including various fault types and severities, as described in Table 3. Vibration signals are split into sections with equal window length (1024 data points). Based on RQA technique, for visualization, the corresponding RPs and feature distribution are shown in Fig. 3, where each RP is corresponding to one sample of 1024 points, and the box plot is corresponding to the distribution of normalized features. Most of the samples can be distinguished from others by the distribution of normalized feature, or the texture feature of different RPs. However, when B (0.014) or OR (0.021) occurs, their RPs are too similar and indistinguishable, besides, their feature distributions are also dispersed, just as in Fig. 3 c and j. Additionally, it can be noticed that F9, F10, and F11 are sensitive to bearing health states by contrasting those distribution box plots. Once abnormal bearing conditions occur, their distribution centers change from 1 to other values.

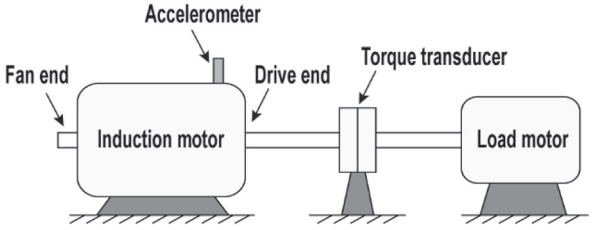

Fig. 2. Rolling bearing test rig.

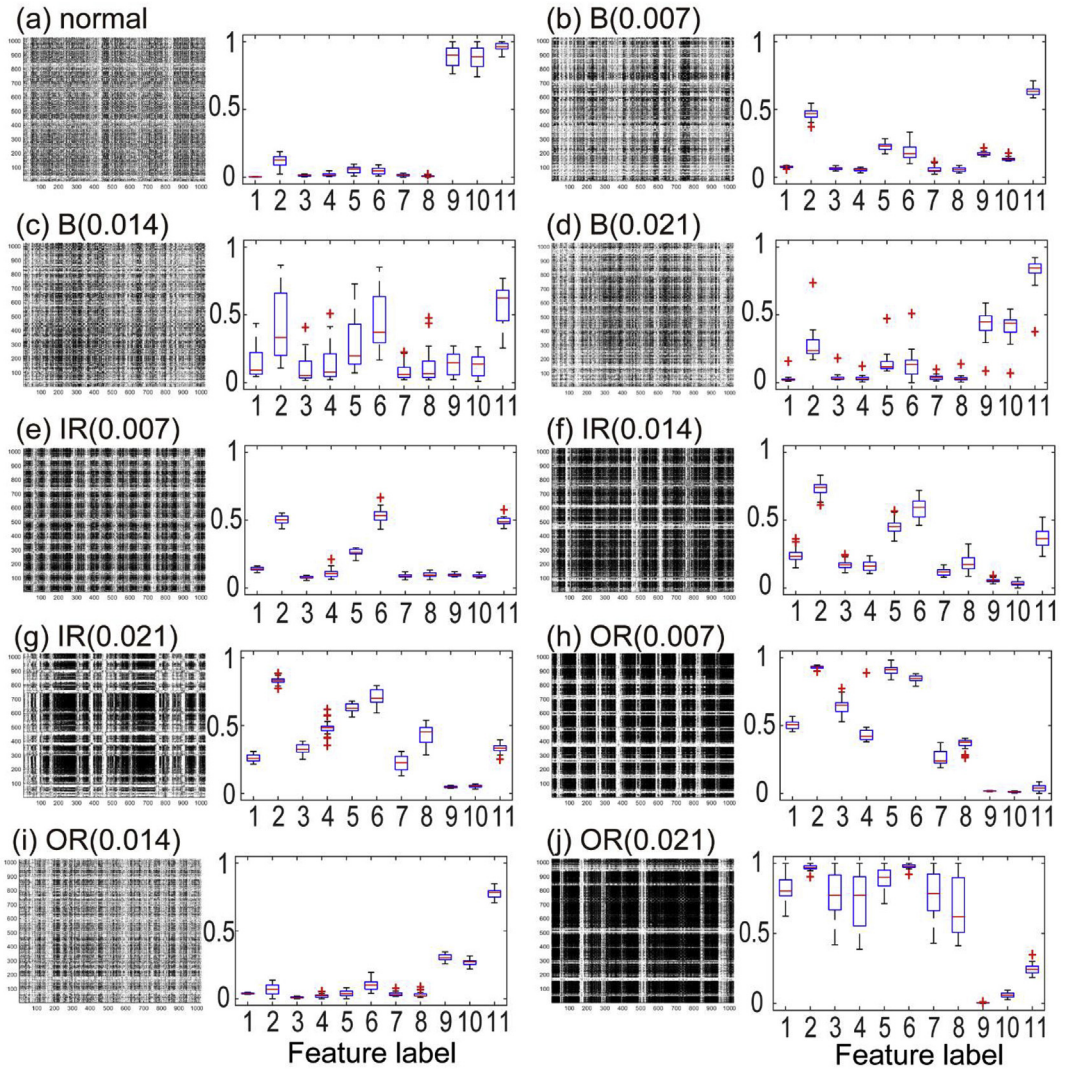

Fig. 3. Box plot of 11 RQA parameters and corresponding RPs: (a) Normal condition, (b) B (0.007), (c) B (0.014), (d) B (0.021), (e) IR (0.007), (f) IR (0.014), (g) IR (0.021), (h) OR (0.007), (i) OR (0.014), (j) OR (0.021).

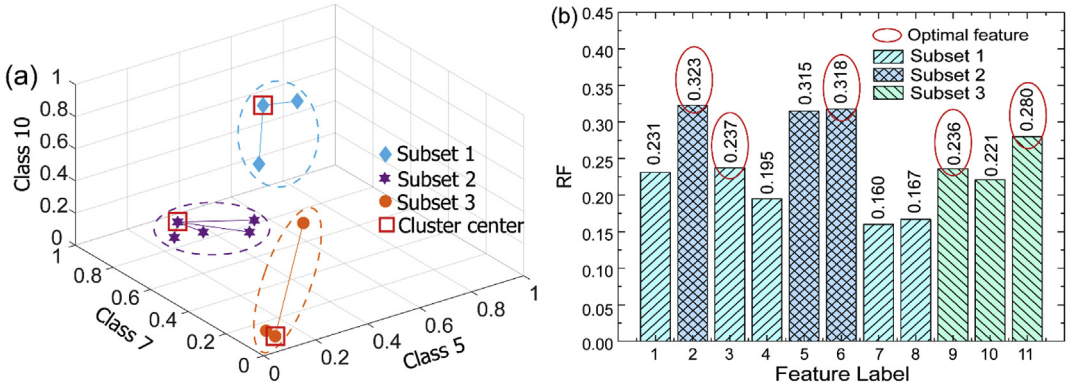

Fig. 4. Scatter plot of feature clustering results (a) and their RFs (b).

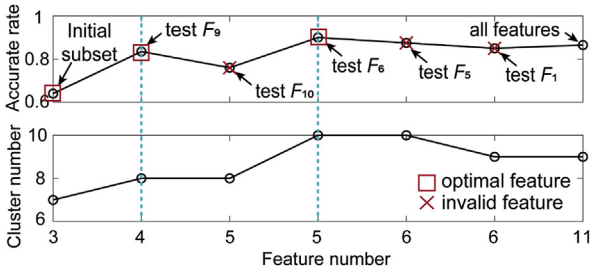

Fig. 5. Visualization flowchart of ARSFS.

Table 4
AP classification accuracy with three feature selection methods.

Methods	Feature sets	Abbreviation	Accuracy
None	Raw features $\{F_1 - F_{11}\}$	11 RFs	86.5%
Proposed method	Optimal features $\{F_2, F_3, F_6, F_9, F_{11}\}$	5 OFs	90%
AP cluster	Center features $\{F_1, F_6, F_9\}$	3 CFs	60%
RReliefF-SFS	Sensitive features $\{F_2, F_3, F_5, F_6, F_9, F_{11}\}$	6 SFs	87.5%

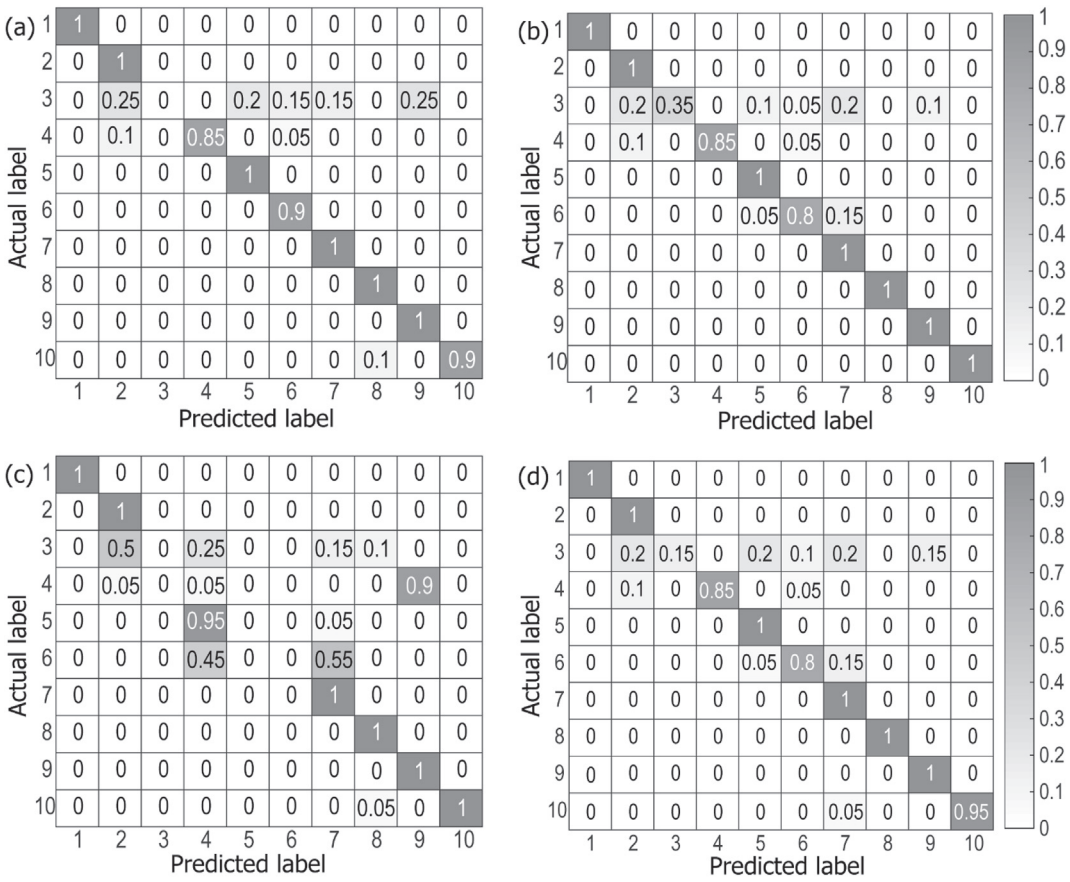

Fig. 6. Scatter plots of multi-class confusion matrix by different feature selection methods: (a) Raw features, (b) ARSFS, (c) AP, (d) RReliefF-SFS.

5.2. Feature selection

As shown in Fig. 3, some features do contribute to the identification of fault categories, but some features are too discrete in distribution, resulting in a fuzzy feature correlation. In the process of feature selection (ARSFS method), as can be seen from

Fig. 4a, 11 raw features are grouped into three subsets, where the RFs of features in the same subset are similar (in Fig. 4b), which well demonstrates the effectiveness of the AP method to cluster homogenous features, as mentioned in Section 2.

Homogeneous features may have similar contributions to pattern identification, and in many cases, their presence might bring multiple-correlations into the feature space. For visualization of ARSFS, the variation of clustering accuracy and the number of categories in the process are both presented in Fig. 5, where the actual number of sample classes is 10. When $\tilde{\mathbf{S}} = \{F_2, F_3, F_6, F_9, F_{11}\}$, the highest accuracy is achieved. To sum up, according to the proposed feature selection loop, 5 key features are selected automatically, as shown in Fig. 4b.

To further verify the performance of ARSFS method, the feature center method and RReliefF-SFS [33] are introduced as contrast methods. Refer to [13], AP clustering is used to find and extract the feature centers, besides the results are listed in Table 4. In addition, 11 raw features (11 RFs), 5 optimal features (5 OFs using ARSFS), 3 center features (3 CFs using AP clustering), and 6 sensitive features (6 SFs) are all applied in an unsupervised classifier (AP classifier) to observe the sample distribution in different feature spaces.

From Table 4, the test accuracy of the AP clustering for each feature set is 86.5% (11 RFs), 90% (5 OFs), 60% (3 CFs) and 87.5% (6 SFs), respectively. Fig. 6 is the multi-class confusion matrix of the AP clustering for different each feature set. The multi-class confusion matrix is a visual record of the classification results of bearing in different states, including the correct classification information and the distribution of error classification. The vertical axis of a multi-class confusion matrix represents the actual label for each bearing state, and the horizontal axis represents the predicted label. Therefore, the diagonal elements of multi-class confusion matrix represent the classification accuracy of each condition. As can be seen from Fig. 6, although the third bearing state is the hardest to be distinguished, the overall accuracy of the recognition has been significantly improved after ARSFS feature selection. It is much higher than other feature selection methods (in Fig. 6c–d), which are 60% and 87.5%, respectively.

5.3. Fault identification performance

The selected five features $\{F_2, F_3, F_6, F_9, F_{11}\}$ are composed into a feature vector matrix in fault diagnosis of the rolling bearing. All 900 samples are split into a training set compromising 600 samples and a test set compromising 300 samples without overlapping. For a 10-class problem, the results in Table 5 demonstrate the AOR-VPMCD classification process, where the test sample label equals the subscript k of the predictive model AOR_k with the least SSE. In most cases, it can be considered as a feasible approach.

Based on different feature spaces, such as 5 OFs, 11 RFs, and 5 multiple linear correlative features (5 feature in subset 2), the classification accuracy of different VPMs (listed in Table 6) would help to analyze the applicable range of the two models,

Table 5
Results of pattern recognition of rolling bearing based on AOR-VPMCD.

Actual Label	SSE										Predicted Label
	AOR_1	AOR_2	AOR_3	AOR_4	AOR_5	AOR_6	AOR_7	AOR_8	AOR_9	AOR_{10}	
1	**1.1e-4**	7.8e7	1.4e5	9.4e3	4.2e7	2.0e7	4.9e8	2.8e9	4.1e4	2.8e9	**1**
2	2.2e5	**2.7e-3**	111.8	45.0	194.3	1.6e3	1.4e5	7.9e7	92.9	1.7e8	**2**
3	6.6e4	1.5e3	**7.8e-2**	1.3e3	371.7	5.7e5	4.1e6	1.6e8	3.3	2.3e8	**3**
4	217.9	5.0e5	339.2	**7.9e-3**	9.7e5	1.4e6	5.2e7	1.0e9	1.2e3	1.2e9	**4**
5	6.5e6	1.2e3	11.5	8.3e3	**1.7e-3**	103.1	9.4e3	1.5e6	5.8e3	2.7e6	**5**
6	2.4e7	1.5e3	25.2	1.2e3	90.8	**9.1e-4**	131.6	1.2e5	1.2e5	3.3e5	**6**
7	9.2e7	1.5e5	2.3e3	4.2e5	4.2e3	51.1	**9.5e-3**	8.7e3	5.4e5	2.73	**7**
8	1.9e8	1.0e7	1.1e5	2.5e7	1.3e6	4.2e3	512.6	**2.5e-4**	5.2e6	569.4	**8**
9	4.1e4	4.2e4	25.1	2.9e4	1.1e5	1.4e7	7.9e7	1.2e9	**3.4e-3**	1.7e9	**9**
10	3.3e8	1.1e7	1.7e4	5.5e7	1.0e6	2.8e3	118.7	1.0e3	4.8e6	**6.5e-4**	**10**

Table 6
Comparative results of fault identification accuracy with different feature set.

Feature Set	Accuracy (%)						
	Training number per state	10	20	30	40	50	60
5 OFs	GRNN-VPMCD	90.11	91.33	91.27	91.50	91.47	91.40
	RBF-VPMCD	83.54	86.96	94.84	96.67	97.33	98.02
	μ	1	1	2	2	2	2
5 MFs	GRNN-VPMCD	81.66	81.71	81.76	81.96	82.16	82.15
	RBF-VPMCD	71.55	82.88	88.29	89.96	88.86	89.50
	μ	1	2	2	2	2	2
11 RFs	GRNN-VPMCD	91.70	91.88	92.62	92.66	93.06	93.33
	RBF-VPMCD	79.76	90.70	90.26	91.51	91.46	91.47
	μ	1	1	1	1	1	1

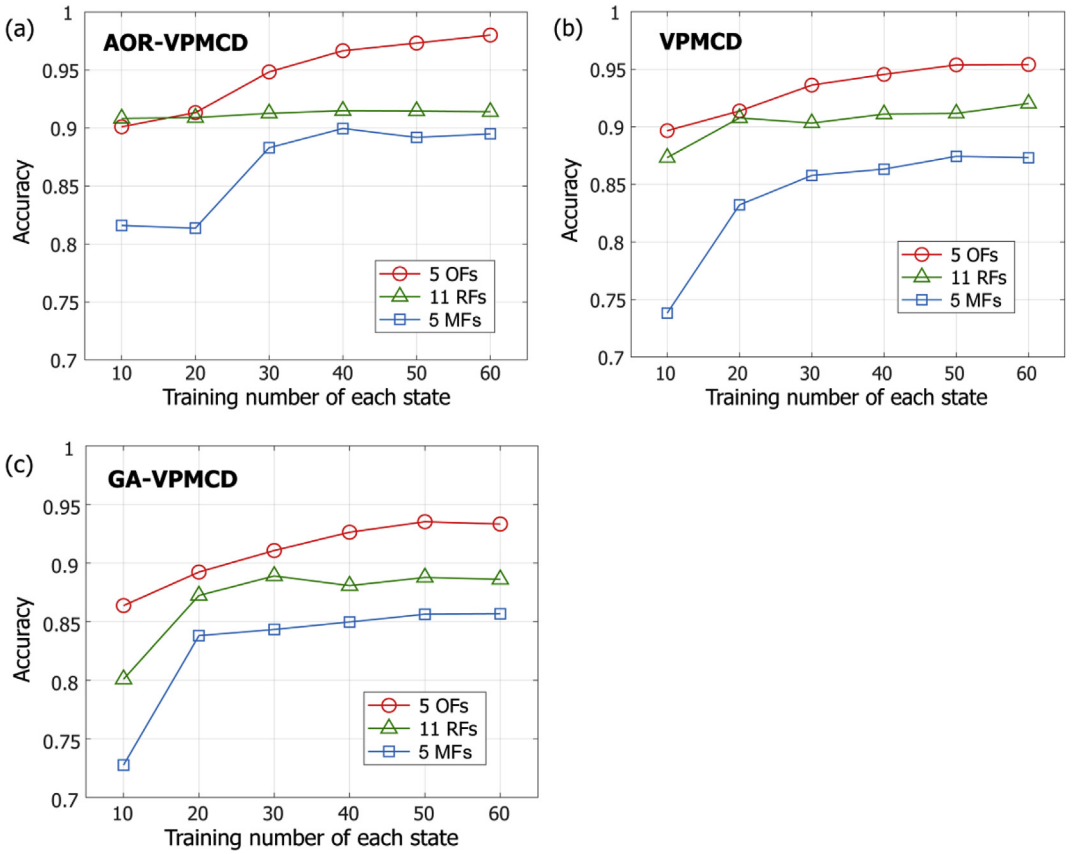

Fig. 7. Recognition result based on different classifiers and different feature sets: (a) AOR-VPMCD, (b) VPMCD, (c) GA-VPMCD.

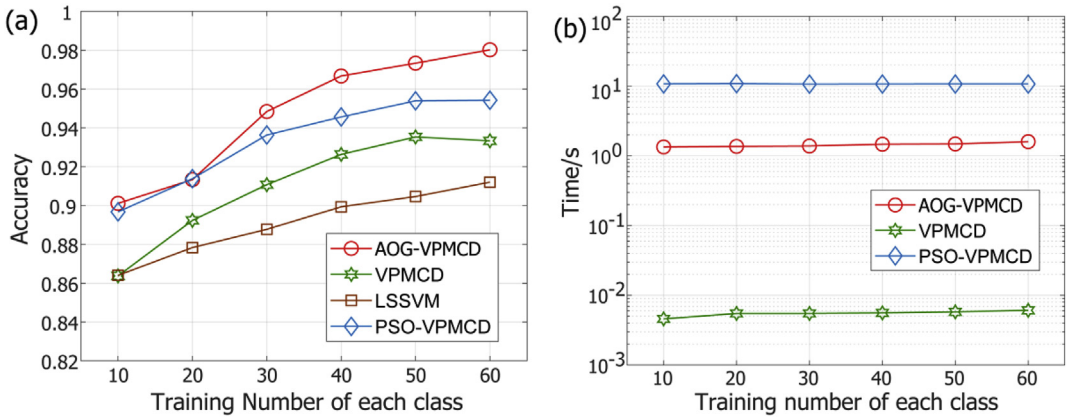

Fig. 8. Performance comparison with different algorithms: (a) accurate rate comparison, (b) Time consumption comparison.

namely RBF-VPM and GRNN-VPM. Furthermore, the number of training sample varies from 10 to 60 per class to analyze the influence of the number.

Firstly, based on the optimal feature set (5 OFs). When the number of training sample is less than 30 per class, the experimental results show that GRNN-VPMCD has good robustness to small samples. Secondly, the best accuracy (98.02%) is achieved by RBF-VPMCD with 5 OFs and sufficient training sample. Thirdly, owing to the influence of multiple linear correlations in 5 MFs on VPMCD, the overall recognition accuracy is reduced by 10%. Finally, based on the original feature set, unlike the first two cases, GRNN-VPMCD performance is always superior to the RBF-VPMCD, where the parameter μ is always

set to 1. To sum up, aiming at redundancies in the feature space or insufficient training samples, GRNN-VPMCD is a good choice.as the output of it converges to the optimal regression surface. However, its performance is easy to reach a plateau, and then it is difficult to make further improvements. In addition, after fully training, the performance of RBF-VPMCD is obviously outstanding.

In terms of feature selection methods, we use the same training and testing samples to analyze the performance of AOR-VPMCD, VPMCD, and GA-VPMCD. Fig. 7 shows the results of the recognition rate, indicating that the application of the proposed ARSFS before the classifier can further improve its performance. For example, AOR-VPMCD has an average increase of 5% in accuracy compared to not having a feature selection process. In addition, Fig. 7a shows that the average performance of AOR-VPMCD is better than that of VPMCD and GA-VPMCD, using 5 multiple linear correlative features. It is noteworthy that, even for small sample data, AOR-VPMCD still maintains good reliability. And when there are multiple linear correlation features (5 MFs), the accuracy of AOR-VPMCD can still reach 90%. The result of classification not only validates the robustness and practicability of AOR-VPMCD, but also verifies the contribution of the proposed feature selection method.

Finally, the VPMCD, GA-VPMCD, and LSSVM based on ARSFS method are compared with each other. Fig. 8 shows a comparison of the accuracy of the classification and the time required. Although the accurate rate increased with the number of training samples, the performance of VPMCD (86.37%–93.33%) and GA-VPMCD (89.67%–95.42%) tended to be saturated with the increasing training samples. At the same time, the accuracy of LSSVM (86.4%–91.2%) is always lower than average. On the contrary, AOR-VPMCD is always the best. As far as the efficiency is concerned, it can be observed from Fig. 8b that although VPMCD is almost instantaneous, AOR-VPMCD (average 1.44s) is also much better than GA-VPMCD (average 10.73s).

6. Conclusion

The result shows that the proposed feature selection and intelligent classification method both have some potential applications in practical. Particularly, compared with VPMCD and its various improved methods. The research contents of this paper include: (1) A new feature selection method is proposed for the redundancy and multiple linear relationships in the original feature space. (2) Considering the good non-linear fitting capability of RBF and GRNN, two new kinds of VPMs are defined. (3) An entire fault diagnosis process is proposed and applied to the identification of various rolling bearing states.

In addition, there are still some problems on AOR-VPMCD. Firstly, the diffusion factor ξ and width factor σ_n are both artificial parameters belonging to neural network, and their values are related to the state of the training sample. Secondly, more experiments are needed to find a clear cut-in point between RBF-VPM and GRNN-VPM. Therefore, how to select the best parameters adaptively for each AOR-VPM and how to define the scope and boundaries of each model are still needed to be solved in the future work.

Acknowledgements

This work is supported by the projects from the National Science Foundation of China (No. 51475052, 51675064), the Fundamental Research Funds for the Central Universities (No. 106112016CDJZR115502), China Postdoctoral Science Foundation (No. 2016T90833 and 2015M582519), Chongqing Postdoctoral Science Foundation (No. Xm2016018).

References

[1] Y. Liu, B. He, F. Liu, S. Lu, Y. Zhao, Feature fusion using kernel joint approximate diagonalization of eigen-matrices for rolling bearing fault identification, J. Sound Vib. 385 (2016) 389–401, https://doi.org/10.1016/j.jsv.2016.09.018.
[2] L. Xie, M. Yin, Q. Huang, Y. Zhao, Z. Deng, Z. Xiang, G. Yin, Internal defect inspection in magnetic tile by using acoustic resonance technology, J. Sound Vib. 383 (2016) 108–123, https://doi.org/10.1016/j.jsv.2016.07.020.
[3] R. Raghuraj, S. Lakshminarayanan, VPMCD: variable interaction modeling approach for class discrimination in biological systems, FEBS Lett. 581 (2007) 826–830, https://doi.org/10.1016/j.febslet.2007.01.052.
[4] B.D. Barkana, I. Saricicek, B. Yildirim, Performance analysis of descriptive statistical features in retinal vessel segmentation via fuzzy logic, ANN, SVM, and classifier fusion, Knowl. Based Syst. 118 (2017) 165–176, https://doi.org/10.1016/j.knosys.2016.11.022.
[5] C.A.D.A. Padilha, D.A.C. Barone, A.D.D. Neto, A multi-level approach using genetic algorithms in an ensemble of Least Squares Support Vector Machines, Knowl. Based Syst. 106 (2016) 85–95, https://doi.org/10.1016/j.knosys.2016.05.033.
[6] Y. Yang, H. Pan, L. Ma, J. Cheng, A roller bearing fault diagnosis method based on the improved ITD and RRVPMCD, Measurement 55 (2014) 255–264, https://doi.org/10.1016/j.measurement.2014.05.016.
[7] S. Luo, J. Cheng, K. Wei, A fault diagnosis model based on LCD-SVD-ANN-MIV and VPMCD for rotating machinery, Shock Vib. 2016 (2016), https://doi.org/10.1155/2016/5141564.
[8] L. Breiman, Heuristics of instability and stabilization in model selection, Ann. Stat. 24 (1996) 2350–2383, https://doi.org/10.1214/aos/1032181158.
[9] H. Cui, M. Hong, Y. Qiao, Y. Yin, 2308. Application of VPMCD method based on PLS for rolling bearing fault diagnosis, J. Vibroeng. 19 (2017) 160–175. http://10.0.84.91/jve.2016.17156.
[10] S. Luo, J. Cheng, M. Zeng, Y. Yang, An intelligent fault diagnosis model for rotating machinery based on multi-scale higher order singular spectrum analysis and GA-VPMCD, Meas. J. Int. Meas. Confed. 87 (2016) 38–50, https://doi.org/10.1016/j.measurement.2016.01.006.
[11] F. Charte, A.J. Rivera, M.J. Del Jesus, F. Herrera, MLSMOTE: approaching imbalanced multilabel learning through synthetic instance generation, Knowl. Based Syst. 89 (2015) 385–397, https://doi.org/10.1016/j.knosys.2015.07.019.
[12] J. Song, C.E. Romero, Z. Yao, B. He, A globally enhanced general regression neural network for on-line multiple emissions prediction of utility boiler, Knowl. Based Syst. 118 (2017) 4–14, https://doi.org/10.1016/j.knosys.2016.11.003.
[13] Z. Wei, Y. Wang, S. He, J. Bao, A novel intelligent method for bearing fault diagnosis based on affinity propagation clustering and adaptive feature selection, Knowl. Based Syst. 116 (2017) 1–12, https://doi.org/10.1016/j.knosys.2016.10.022.
[14] V. Abrahamsson, N. Ristic, K. Franz, K. Van Geem, Comprehensive two-dimensional gas chromatography in combination with pixel-based analysis for fouling tendency prediction, J. Chromatogr. A 1501 (2017) 89–98, https://doi.org/10.1016/j.chroma.2017.04.021.

[15] M. Al-Rajab, J. Lu, Q. Xu, Examining applying high performance genetic data feature selection and classification algorithms for colon cancer diagnosis, Comput. Meth. Progr. Biomed. 146 (2017) 11–24, https://doi.org/10.1016/j.cmpb.2017.05.001.
[16] B.J. Frey, D. Dueck, Clustering by passing messages between data points, Science (80-.) 315 (2007) 972–976, https://doi.org/10.1126/science.1136800.
[17] I. Kononenko, Estimating attributes: analysis and extensions of RELIEF, in: Eur. Conf. Mach. Learn, Springer, 1994, pp. 171–182.
[18] X. Liu, L. Bo, Identification of resonance states of rotor-bearing system using RQA and optimal binary tree SVM, Neurocomputing 152 (2015) 36–44, https://doi.org/10.1016/j.neucom.2014.11.021.
[19] Y. Qian, R. Yan, S. Hu, Bearing degradation evaluation using recurrence quantification analysis and kalman filter, IEEE Trans. Instrum. Meas. 63 (2014) 2599–2610, https://doi.org/10.1109/TIM.2014.2313034.
[20] N. Marwan, J. Kurths, Nonlinear analysis of bivariate data with cross recurrence plots, Phys. Lett. Sect. A Gen. At. Solid State Phys. 302 (2002) 299–307, https://doi.org/10.1016/S0375-9601(02)01170-2.
[21] J.-P. Eckmann, S.O. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems, Europhys. Lett. 4 (1987) 973–977, https://doi.org/10.1209/0295-5075/4/9/004.
[22] H.D.I. Abarbanel, N. Masuda, M.I. Rabinovich, E. Tumer, Distribution of mutual information, Phys. Lett. Sect. A Gen. At. Solid State Phys. 281 (2001) 368–373, https://doi.org/10.1016/S0375-9601(01)00128-1.
[23] L. Cao, Practical method for determining the minimum embedding dimension of a scalar time series, Phys. D Nonlinear Phenom. 110 (1997) 43–50, https://doi.org/10.1016/S0167-2789(97)00118-8.
[24] G. Gan, M.K.-P. Ng, Subspace clustering using affinity propagation, Pattern Recogn. 48 (2015) 1455–1464, https://doi.org/10.1016/j.patcog.2014.11.003.
[25] F. Shang, L.C. Jiao, J. Shi, F. Wang, M. Gong, Fast affinity propagation clustering: a multilevel approach, Pattern Recogn. 45 (2012) 474–486, https://doi.org/10.1016/j.patcog.2011.04.032.
[26] M. Robnik-Šikonja, I. Kononenko, Theoretical and empirical analysis of ReliefF and RReliefF, Mach. Learn. 53 (2003) 23–69, https://doi.org/10.1023/A:1025667309714.
[27] R.R. Da Silva, M.H.S. Siqueira, L.P. Calôba, I.C. Da Silva, A. De Carvalho, J.M. Rebello, Contribution to the development of a radiographic inspection automated system, J. Nondestruct. Test 7 (2002) 1–8.
[28] A.K. Jain, R.P.W. Duin, J. Mao, Statistical pattern recognition: a review, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 4–37, https://doi.org/10.1109/34.824819.
[29] I.R. Donis-González, D.E. Guyer, A. Pease, Postharvest noninvasive classification of tough-fibrous asparagus using computed tomography images, Postharvest Biol. Technol. 121 (2016) 27–35, https://doi.org/10.1016/j.postharvbio.2016.07.012.
[30] I. Mandal, Developing new machine learning ensembles for quality spine diagnosis, Knowl. Based Syst. 73 (2014) 298–310, https://doi.org/10.1016/j.knosys.2014.10.012.
[31] H. Li, S. Guo, C. Li, J. Sun, A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm, Knowl. Based Syst. 37 (2013) 378–387, https://doi.org/10.1016/j.knosys.2012.08.015.
[32] W. Shen, X. Guo, C. Wu, D. Wu, Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm, Knowl. Based Syst. 24 (2011) 378–385, https://doi.org/10.1016/j.knosys.2010.11.001.
[33] M. Wu, Y. Wang, A feature selection algorithm of music genre classification based on ReliefF and SFS, in: Comput. Inf. Sci. (ICIS), 2015 IEEE/ACIS 14th Int. Conf., IEEE, 2015, pp. 539–544.