Measurement 133 (2019) 468–478

Contents lists available at ScienceDirect

Measurement

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

A new feature extraction approach using improved symbolic aggregate approximation for machinery intelligent diagnosis

Yulong Zhang [a], Lixiang Duan [b,*], Menglan Duan [a]

[a] Institute for Ocean Engineering, China University of Petroleum, Beijing 102249, China
[b] School of Mechanical and Transportation Engineering, China University of Petroleum, Beijing 102249, China

ARTICLE INFO

Article history:
Received 7 August 2017
Received in revised form 3 May 2018
Accepted 15 October 2018
Available online 17 October 2018

Keywords:
Symbolic Aggregate approXimation (SAX)
Improved SAX
Bearing defect diagnosis
Reciprocating compressor valve
Fault pattern

ABSTRACT

Feature extraction from vibration signals is considerably significant for condition monitoring and fault diagnosis. The Symbolic Aggregate approXimation (SAX) technique essentially transforming a real-valued time series into a symbol sequence, has been proven as a potential tool of feature extraction for machinery intelligent diagnosis. However, challenge still exists that the SAX cannot fulfill feature extraction tasks well since it is carried out only on the basis of mean value in time domain. To overcome this limitation, an improved SAX (ISAX) is proposed in this paper. This new method substitutes the feature of mean value in time domain with multiple features extracted from time, frequency and time-frequency domains in order to obtain comprehensive fault information. With the ISAX transformation, a vibration signal can be transformed into various symbol sequences according to the multi-domain features. Next the Shannon entropy technique is conducted on a symbol sequence to capture sequential patterns in local signals and then the Shannon entropy value is used as the eigenvalue of the symbol sequence. Various eigenvalues are obtained to describe a vibration signal from different perspectives, which leads to a better feature extraction. These eigenvalues are then fed into the Kernel Principal Component Analysis (KPCA) to reduce dimensions and extract principal features for classification tasks. Compared with SAX, the most significant advantage of ISAX is extracting comprehensive signal characteristics from multi-domain. Moreover, the ISAX captures fault information better considering the local fault patterns. The effectiveness and superiority of ISAX were validated by experimental studies using the fault signals of rolling bearings and reciprocating compressor valves with remarkably high classification rates.

1. Introduction

The rapid advancements in modern manufacturing aimed to achieve higher productivity, better quality, and more flexibility depend on not only the innovations in information and industry fields, but also the developments of monitoring and health diagnosis techniques [1]. The growing demands on fault-free operations of various components in manufacturing machines accelerate progress in sensing techniques of machinery condition monitoring, such as vibration analysis [2], oil analysis [3], acoustic emission [4], pressure [5], infrared thermography [6] and other nondestructive testing techniques [7].

Vibration signals have been extensively used for machinery fault diagnosis due to the convenience of installing sensors and collecting condition information without interruption of running conditions [8]. Numerous efforts have been made to develop feature extraction techniques based on vibration signals and these methods can be categorized in time domain [9,10], frequency domain [11–13] and time-frequency domain [14–16].

Time domain methods generally give information of signal amplitude varying with time but the time domain features are unable to detect faults at early stage [17]. Frequency domain methods are alternative techniques for feature extraction from the perspective of frequency analysis. Every mechanical component has its own characteristic frequency, so some defects can be identified with the use of frequency analysis methods. Yet disadvantages of frequency domain analysis methods lie in their poor abilities to analyze the non-stationary signals [18].

For the non-stationary signals, wavelet transform (WT) which is a classical analysis method in time-frequency domain, allows the signals to be decomposed into many different frequency bandwidths to extract fault features. It has been widely investigated due to its strict mathematical foundation, high time-frequency

* Corresponding author.
E-mail address: duanlx@cup.edu.cn (L. Duan).

https://doi.org/10.1016/j.measurement.2018.10.045
0263-2241/

resolution and multi-scale analysis features [19]. However, the selection of mother wavelet functions and wavelet decomposition levels is usually determined with experience, which could result in great subjective influence on results. Different from WT, the empirical mode decomposition (EMD) [20] was proposed as a self-adaptive algorithm, which can adaptively decompose the signals into intrinsic mode functions (IMFs) to reflect the elaborate time-frequency energy distribution of the signals. It has become a newly developed powerful method for non-linear and non-stationary time series analysis in signal processing and other engineering fields [21–23]. Unfortunately, there also exists drawbacks with this algorithm, such as the fitting errors caused by cubic splines, mode mixing, end effects and the negative frequency caused by Hilbert transform [19].

Recently, an EMD-like adaptive decomposition method namely empirical wavelet transform (EWT) was proposed by Gilles for non-stationary signal processing and it has attracted growing interests [24]. This new method integrates the advantages of the EMD and WT. Some applications of this technique in the area of machinery fault diagnosis have been reported. He et al employed EWT for wheel-bearing fault diagnosis of trains [25]. Kedadouche et al proposed a new approach based on OMA-EWT for bearing fault diagnosis [26]. Chen et al modified EWT via data-driven adaptive Fourier spectrum segment for mono-component feature extraction [27] and employed EWT for generator bearing fault diagnosis of wind turbine [28].

It is clear that a variety of feature extraction techniques have been carried out in the field of condition monitoring. In recent years, the research of symbolization has been reported a lot in the data mining field, but it has been little investigated for fault diagnosis. The basic idea of symbolization is transforming a continuous valued time series into a discrete representation, giving a much higher computation efficiency via dimensionality reduction [29]. This fast execution is crucial for online computation involved in huge amounts of data.

Along this path, George Georgoulas introduced SAX symbolic representation to transform vibration signals into a discrete valued sequence and an intelligent icon for defect diagnosis of rolling bearings [30,31]. The original SAX technique depends on the Piecewise Aggregate Approximation (PAA) representation for dimensionality reduction, which substitutes the mean values of equal sized frames in time domain with symbol representation. However, such mean value based representation in time domain has a poor ability to extract features because it may miss the information in frequency domain and time-frequency domain.

To address the aforementioned issues, an improved SAX is proposed to map vibration signals into symbol sequences according to multi-domain features for machinery intelligent diagnosis. To obtain comprehensive condition information, the original SAX algorithm is improved by substituting the mean value of equal sized frames in time domain with multiple features extracted from time, frequency and time-frequency domains. With the ISAX transformation, a vibration signal can be transformed into various symbol sequences according to the multi-domain features. Next the Shannon entropy technique is conducted on a symbol sequence to capture sequential patterns in local signals and then the Shannon entropy value is used as the eigenvalue of the symbol sequence. Various eigenvalues from different perspectives can be obtained and these eigenvalues are then fed into a data fusion tool to create representative features for classification tasks. The experimental studies on rolling bearings and reciprocating compressor valves suggested that the proposed method is highly effective for the condition monitoring and fault identification.

The intellectual merits of this paper rest on two folds: (1) a novel symbol representation approach for feature extraction is presented in the field of fault diagnosis and this new proposition focuses on multi-domain signal characteristics and meanwhile provides a high-efficiency computation especially for online fault identification with large volumes of data; and (2) the fault pattern is first investigated to reflect the local fault information by means of the Shannon entropy theory, leading to a better feature extraction. The remainder of this paper is organized as follows. The principle of SAX is introduced firstly in Section 1 and the improvements of SAX are then investigated by extending into multi-domain in Section 2. Next, the feature extraction approach using ISAX is described and then a fault diagnosis framework is proposed in Section 3 followed by experimental verification in Section 4. Then the effectiveness of the proposed method is demonstrated in Section 5. Lastly, conclusions are drawn in Section 6.

2. Methodology

SAX algorithm is a technique of symbolic representation of time series, it transforms a time series S of length n into a symbol sequence S' of length w, where $w << n$, achieving high-dimensional data compression [28]. This process can be summarized as three steps [33]: (1) A time series is normalized in order to have zero mean and standard deviation of one; (2) The normalized signal is then divided into equal sized sections and the mean value of each section is calculated. By representing each section with its mean value, a process of Piecewise Aggregate Approximation (PAA) for dimension reduction is achieved; (3) After the PAA representation, a symbol discretization process operates in order to produce a sequence with approximately equiprobable symbols. The schematic diagram of SAX process is seen in Fig. 1.

2.1. Symbolic aggregate approximation

2.1.1. Piecewise aggregate approximation

In most cases, the time series $S = \{a_1, a_2, \cdots, a_n\}$ must be normalized to form a new series with a zero mean and a standard deviation of one before transforming into the PAA representation $\bar{S} = \left\{\bar{a}_1, \bar{a}_2, \cdots, \bar{a}_w\right\}$. The normalized series S_1 is:

$$S_1 = \frac{S - \mu}{\sigma} \tag{1}$$

where μ is the mean value of all the elements in S and σ is the standard deviation.

An arbitrary element $\bar{a}_i \in \bar{S}$ can be calculated by:

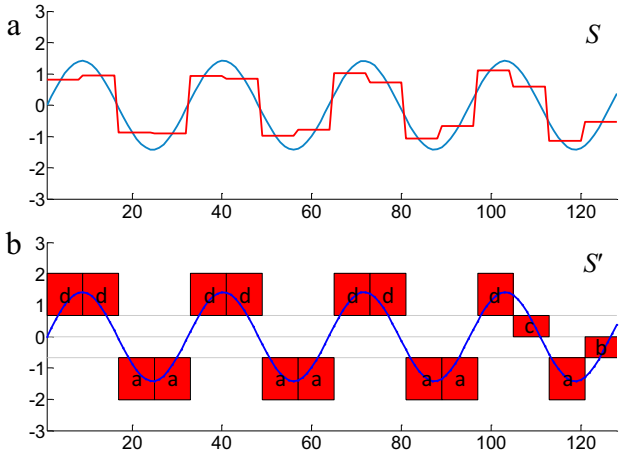

Fig. 1. The SAX transformation: (a) a PAA sequence of length $w = 16$ generated by a time series of length $n = 128$ and (b) the symbol sequence using an alphabet of size four (a, b, c and d) based on the PAA sequence [32].

$$\bar{a}_i = \frac{w}{n} \sum_{j=(i-1)w+1}^{iw} a_j \tag{2}$$

where w and n are the length of the symbol string S' and the time series S_1, respectively. The a_j is an arbitrary element $a_j \in S_1$.

Through the PAA process, long time series can be transformed into a short sequence. The vector of the mean values becomes a new data reduction representation.

2.1.2. Symbol discretization

A further transformation operates in order to obtain a discrete symbol representation after PAA representation. The crucial issue is to make the normalized time series approximately follow a Gaussian distribution by determining "breakpoints" with the aim to produce equal sized areas under a Gaussian curve. These "breakpoints" can be determined by looking up in a statistical table [34].

Once the breakpoints have been determined, all PAA coefficients below the smallest breakpoint are mapped to the symbol "a"; the coefficients greater than or equal to the smallest breakpoint and less than the second smallest breakpoint are mapped to the symbol "b", etc.

2.2. Improved symbolic aggregate approximation

In PAA representation, a time series is divided into k segments with equal length and the mean value in time domain of each segment is employed as a coordinate of a k dimensional feature vector. To improve the SAX, the features extracted from time domain, frequency domain and time-frequency domain are applied as the coordinates in PAA, as seen in Fig. 2. This operation procedure of ISAX is written in Matlab environment. The steps of the algorithm of ISAX is described as Table 1.

3. Formulation of classification model

For a vibration signal, the local sequential varying characteristics of waveform, namely "fault pattern" as seen in Fig. 3(a), can reflect fault information effectively, because different types of defects usually correspond to different fault patterns. Based on this consideration, local structure analysis is proposed to reveal the sequential varying characteristics among neighbored data points and then to identify different fault types. For this purpose, we allow a window with length of l to slide over the symbol sequence to obtain symbol strings, as seen in Fig. 3(b).

3.1. Feature extraction using ISAX

Local structure analysis is employed based on symbol strings to describe the characteristics of fault pattern. Let the alphabet of size $a = 4$, so the symbol set is written as $\phi = (a, b, c, d)$. When the window with length of $l = 2$ slides over the symbol sequence in Fig. 3(b), the set of symbol strings ψ can be written in a summation form as

$$\psi = (bd, da, aa, ab, bc, cb, bd, da, aa, ab, bc, cb, bd, da, aa, ab) \tag{3}$$

When $a = 4$ and $l = 2$, the maximum number of types of symbol strings is $4 \times 4 = 16$. For any symbol sequence composed of an alphabet of size a, the maximum number of types of symbol strings β is given by

$$\beta = a^l \tag{4}$$

where l is the length of symbol string. After the window slides over, the occurrence frequency of one type of symbol string is counted as b_β. The statistical results of occurrence frequency of all types of symbol strings constitute a statistical vector$c$

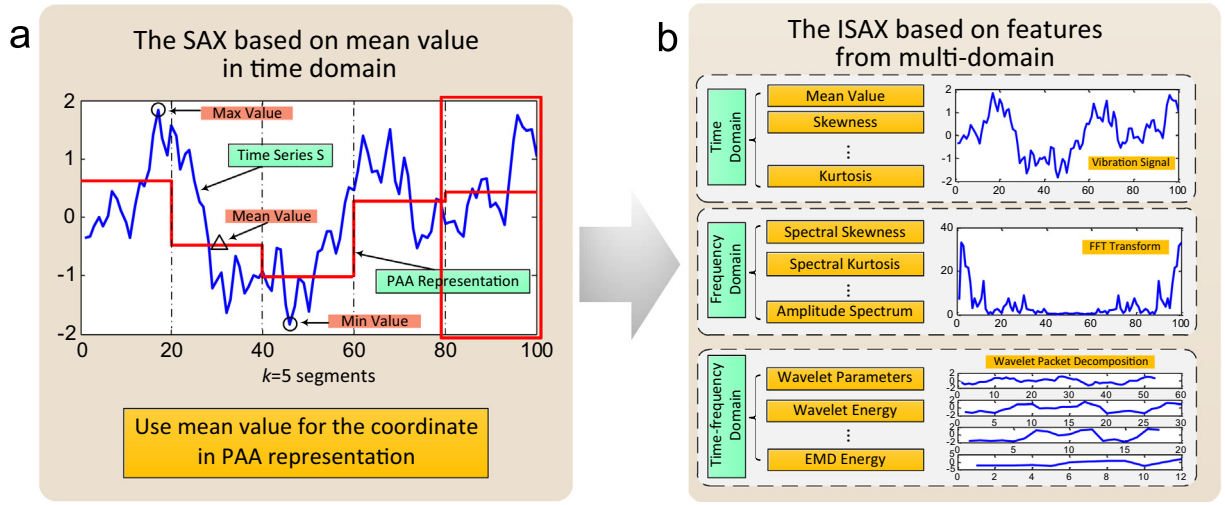

Fig. 2. The scheme of ISAX algorithm: (a) the original SAX uses mean value in time domain for symbol representation, (b) the ISAX uses multi-domain features for symbol representation.

Table 1
Steps of the algorithm of ISAX.

Input: raw vibration signal datasets Output: multi-domain features
(1) Segment the raw vibration signal to obtain data samples. To avoid obtaining abnormal data samples, it is suggested that the length of data sample should be longer than three periodic length of vibration signals (2) In order to eliminate the dimension correlation between variables and form comparable indexes, the data samples need normalization according to Eq. (1) (3) Calculate multi-domain features based on k segments (4) Use the above multi-domain features as coordinates in PAA process for symbol representation, so there are many symbol sequences corresponding to one data sample

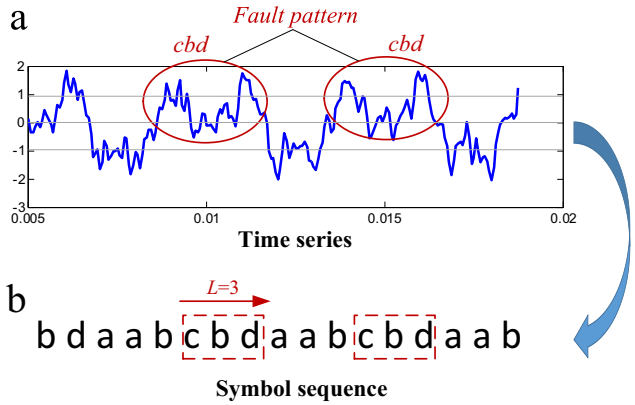

Fig. 3. The flow scheme for "fault pattern" analysis: (a) the fault pattern of "*cbd*" reflecting a typical fault and (b) a slide window with length of $l = 3$ sliding over the symbol sequence to obtain symbol strings.

$$c = (b_1, b_2, \cdots, b_\beta) \tag{5}$$

Again take Fig. 3(b) as an example, b_1 and b_2 can represent the occurrence frequency of "*bd*" and "*da*", respectively.

Then the Shannon entropy algorithm [35] is used for feature extraction. It indicates the irregularity degree of time series and plays a bridging role between signal processing and information theory. The probability p_β of a type of symbol string of length l, can be described as

$$p_\beta = \frac{b_\beta}{n - l + 1} \tag{6}$$

where n is the length of the symbol sequence. According to the statistical vector c, the value of Shannon entropy H is calculated as the eigenvalue of a symbol sequence transformed from the vibration signal sample. The Shannon entropy H of the symbol sequences is summarized as

$$H = -\sum_{\beta=1}^{\alpha^l} p_\beta \lg p_\beta \tag{7}$$

Based on an extracted feature, a symbol sequence is obtained according to Section 2.2 and then an eigenvalue of the symbol sequence is given according to the principle of Shannon entropy. Via the ISAX computation, an original vibration sample can be mapped into various symbol sequences based on multi-domain features and then multiple corresponding eigenvalues can be calculated. These eigenvalues describe the properties of vibration signal from different perspectives. The general procedure of feature extraction using ISAX is summarized as in Fig. 4.

3.2. Diagnosis model

The presented study focuses on the investigation of machinery fault classification using ISAX and discussion on its performance with traditional methods. For this purpose, the acquired vibration signal is firstly segmented into multi-group data samples and then the samples undergo the ISAX process to create various symbol sequences. According to the obtained symbol sequences, a plethora of eigenvalues can be calculated which are viewed as a high-dimension multivariate matrix. It is impractical to input directly this matrix into a training and classification model without dimension reduction. To improve computational efficiency, a proper feature fusion strategy, KPCA, is employed to remove irrelevant and redundant features and lower the dimensions of the feature space [36]. The theoretical framework of KPCA is shown in Appendix A. All the obtained eigenvalues are used as the inputs to the KPCA.

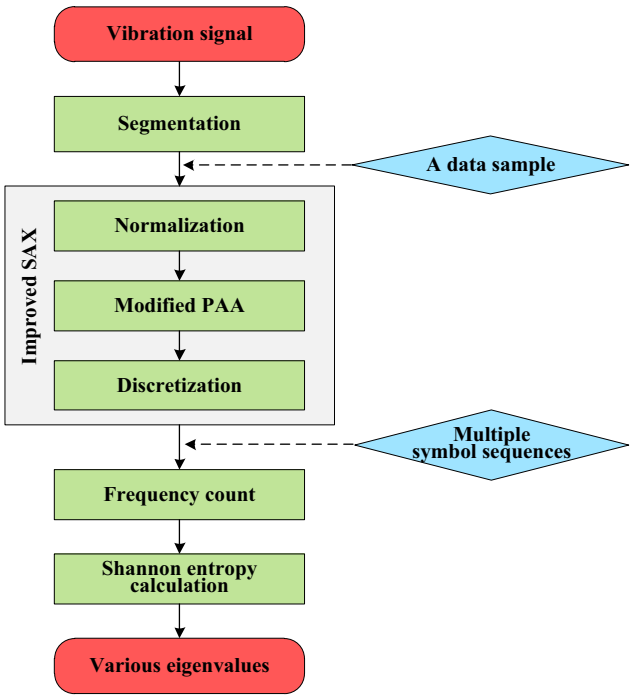

Fig. 4. The scheme of feature extraction based on ISAX.

The number of inputs to the KPCA depends on the number of the extracted multi-domain features. The principal features generated by KPCA are fed into classifiers for defect classification and status evaluation. The general procedure for machinery defect classification is shown in Fig. 5.

4. Experimental studies

To experimentally evaluate the presented method for fault diagnosis, studies were performed on vibration signals measured on bearings in a test rig and reciprocating compressor valves from oil field.

4.1. Diagnosis with experimental data

4.1.1. Vibration datasets

The experimental data come from deep groove ball bearings (6205-2RS JEM SKF) with signal point faults. They were installed in a motor driven mechanical system at the drive end of a motor. Motor bearings were seeded with faults using electro-discharge machining (EDM) to form three types of faults (outer race, inner race and ball faults) ranging from 0.007 to 0.021 in. in diameter. For each test, vibration data was collected using accelerometers placed at the 12o'clock position at the drive end for motor loads of 0–3 horsepower (motor speeds of 1797–1720 RPM) at a sampling frequency of 12,000 samples/s. A more detailed description of the experimental setup and the involved apparatus can be found in Case Western Reserve University's website [37]. To evaluate the proposed approach, a series of experiments using four data subsets (A-D) from the whole data set of the rolling bearings were carried out. The detailed description of the four data subsets was shown in Table 2.

In the data set A, incipient faults were investigated under four different load conditions. The data corresponding to a fault defect size of 0.007 in. irrespective of the location was employed.

In the data set B, the fault data corresponding to a defect size of 0.007 and 0.021 in. was used respectively for training and testing irrespective of the location and the applied loads. The normal data

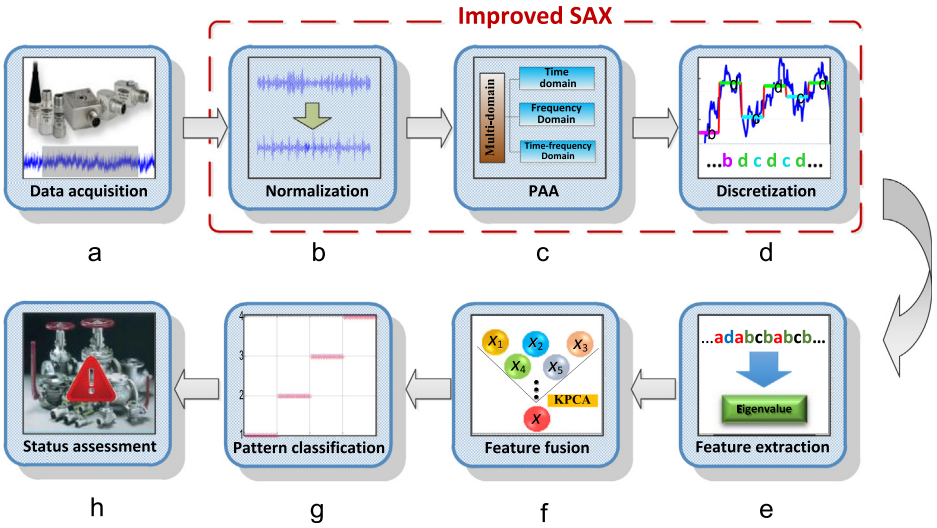

Fig. 5. Illustration of the proposed machinery diagnosis model.

Table 2
Description of the four classification datasets.

Classification settings	Number of training/testing samples	Defect size training/testing (in.)	Conditions	Label
A	400/400	0/0	Normal	1A
	100/100	0.007/0.007	Ball	2A
	150/150	0.007/0.007	Inner race	3A
	300/300	0.007/0.007	Outer race	4A
B	400/–	0/–	Normal	1B
	100/100	0.007/0.021	Ball	2B
	151/150	0.007/0.021	Inner race	3B
	300/301	0.007/0.021	Outer race	4B
C	400/–	0/–	Normal	1C
	100/100	0.021/0.007	Ball	2C
	150/151	0.021/0.007	Inner race	3C
	301/300	0.021/0.007	Outer race	4C
D	400/400	0/0	Normal	1D
	100/100	0.007/0.007	Ball	2D
	150/150	0.007/0.007	Inner race	3D
	300/300	0.007/0.007	Outer race	4D
	100/100	0.014/0.014	Ball	5D
	150/150	0.014/0.014	Inner race	6D
	300/300	0.014/0.014	Outer race	7D
	100/100	0.021/0.021	Ball	8D
	150/150	0.021/0.021	Inner race	9D
	300/300	0.021/0.021	Outer race	10D

was used only for training purpose. The experiments using this data set were carried out to further investigate the classification ability in the case of training by the incipient fault samples.

In the data set C, the same procedure was followed as case B with the exception of the severe fault data for training and the incipient fault data for testing. The purpose of using this data set was to evaluate the classification performance on the incipient faults when the classifier was trained by the serious faulty samples.

In the data set D, 10 subdatasets including three different defect sizes of 0.007, 0.014 and 0.021 in. were employed. The same procedure as in case A was carried out. This experiment was conducted in order to identify the severe grades of faults.

4.1.2. Multi-domain features

A total of 13 features that are commonly encountered in bearing health assessment were extracted from three domains to describe the properties of the vibration signals for bearing defect detection. They were summarized in Table 3.

Ten typical statistical features were extracted from time domain. RMS is an effective measure for the magnitude of a varying quantity and the crest factor is calculated from the peak value divided by the RMS value of the signal. Skewness is used to characterize the degree of signal asymmetry of the distribution around its mean and Kurtosis indicates the spikiness of the signal. Kurtosis coefficient reflects the degree of flat on the top of signal. Features from the frequency domain provide another perspective of bearings operation condition, revealing information that is not easily found in the time domain. In this domain, spectral mean and spectral kurtosis were extracted, where $S(f_i)$ is the power spectrum density. In time-frequency domain, wavelet transform was applied for signal denoising and feature extraction. Features from the time-frequency domain provide another perspective of bearing operation condition, revealing information that is not found in other two domains. The wavelet coefficient with higher energy was selected, which is associated with characteristic frequency in machining. So the energy of selected wavelet coefficient was extracted as a feature.

Table 3
List of extracted features.

Domain	Features	Expression	Label
Time	Mean	$1/n\sum_{i=1}^{n}x_i$	T_1
	Variance	$1/n\sum_{n}(x_i-\bar{x})^2$	T_2
	Maximum	$x \geq x_i (i=1,\cdots,n)$	T_3
	Standard deviation	$\sqrt{\frac{\sum_{i=1}^{n}(x_i-T_{MV})^2}{n-1}}$	T_4
	RMS	$\sqrt{\frac{\sum_{i=1}^{n}x_i^2}{n}}$	T_5
	Root amplitude	$\left(\frac{\sum_{i=1}^{n}\sqrt{x_i}}{n}\right)^2$	T_6
	Crest factor	$\max\frac{\lvert x_i\rvert}{x_{MRS}}$	T_7
	Skewness	$\frac{\sum_{i=1}^{n}x_i^3}{n}$	T_8
	Kurtosis	$\frac{\sum_{i=1}^{n}x_i^4}{n}$	T_9
	Kurtosis coefficient	$\frac{T_K}{T_{RMS}^4}$	T_{10}
'Frequency	Spectral mean	$\sqrt{\sum_{i=1}^{k}f_i^2S(f_i)/\sum_{i=1}^{k}S(f_i)}$	T_{11}
	Spectral kurtosis	$\sum_{i=1}^{K}\left(\frac{f_i-\bar{f}}{\sigma}\right)^4S(f_i)$	T_{12}
Time-frequency	Wavelet energy	$\sum_{i=1}^{K}wt_\varphi^2(i)/N$	T_{13}

These extracted features were used to undergo the ISAX algorithm.

The Shannon entropy values of the symbol sequences in the basis of 13 features over 150 data samples are described schematically in Fig. 6. We can observe that even for one type of fault, the Shannon entropy values of one sample have a significantly difference between others. That is to say, the Shannon entropy value is hypersensitive to slight fluctuations of the raw vibration data and has a good ability of capturing and representing the fault patterns.

For further analysis, the feature of mean value in time domain was selected because it is the coordinate in the original SAX method. The Shannon entropy values based on it over 150 data samples for 4 different conditions were summarized as in Fig. 7. It can be seen that the Shannon values of samples in different conditions may be the same, such as the points A and B, this makes no difference in eigenvalue extraction. So the feature extraction based on the only mean value performs poorly.

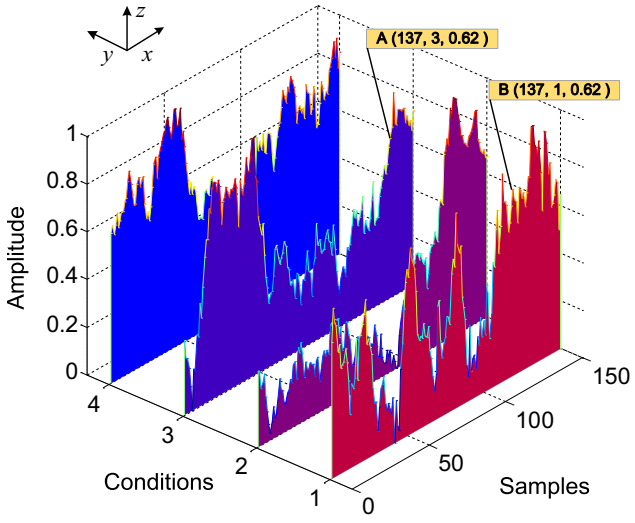

Fig. 7. The Shannon entropy values based on mean value over 150 data samples under 4 different conditions (1-normal; 2-ball fault; 3-inner race fault; 4-outer race fault).

4.2. Diagnosis with field data

Furthermore, experiments were conducted on reciprocating compressor system of model WH64 in a petrochemical plant in northwest China to verify the feasibility of the proposed approach. As illustrated in Fig. 8, it is a 4-cylinder natural gas reciprocating compressor driven by an electric motor with rated power of 1,305 kW. The crankshaft rotation speed is 993 RPM. To collect valve vibration signals for fault diagnosis, an accelerometer was attached to the exhaust valve lid in the 2nd cylinder. A data acquisition system of model MDES-5 designed by China University of Petroleum-Beijing was used for measurements, which envelops a laptop and a data acquisition box configured in a master-slave system. The sampling rate was set to be 16 kHz in this study.

Four sets of vibration signals including normal data and two independent faults (valve wear and valve fracture) and a coupling fault of them (abbreviated N, V_w, S_f, C_f) were used for experiments.

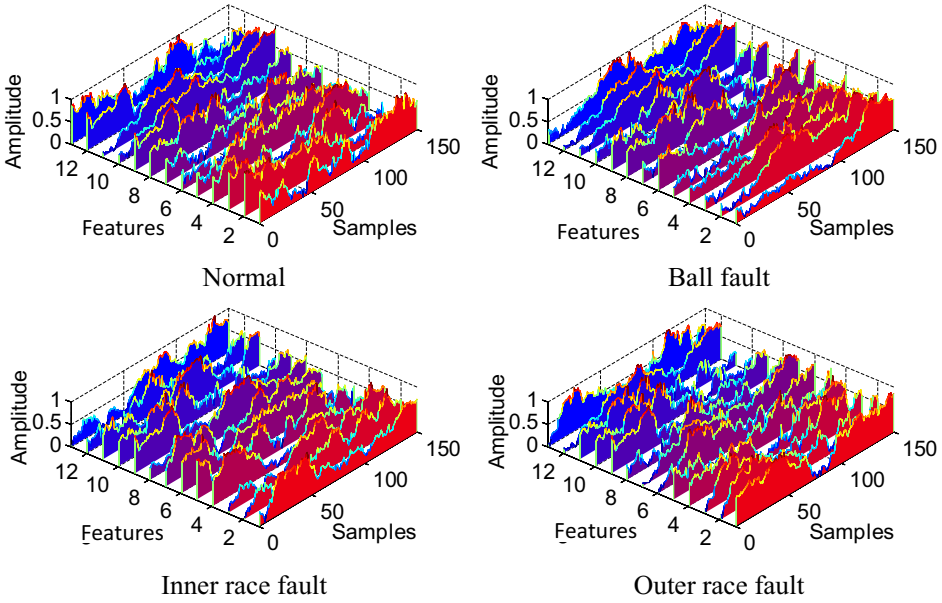

Fig. 6. The Shannon entropy values over 150 data samples with four data sets including normal data and three faults.

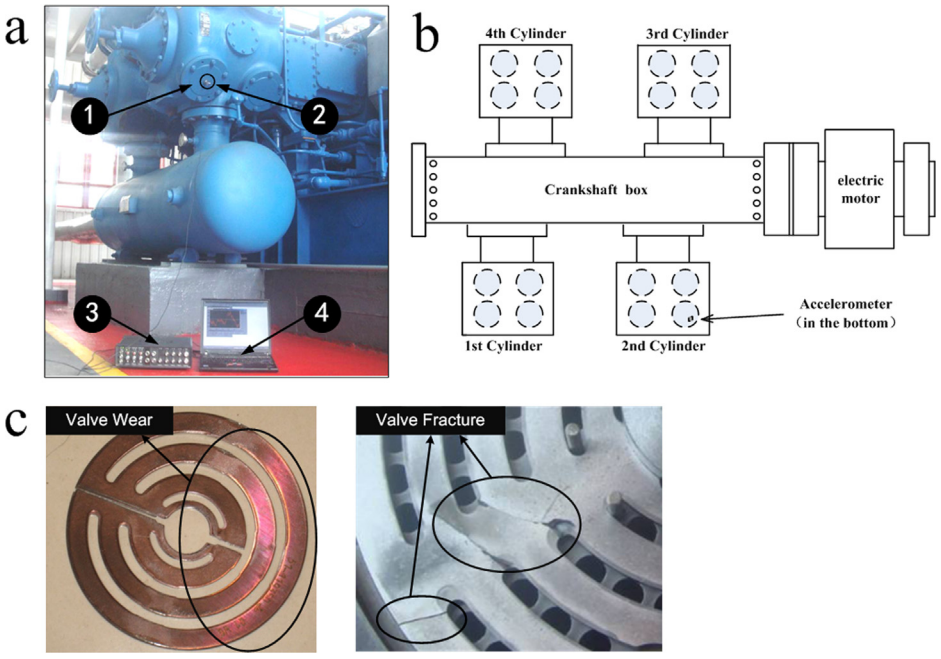

Fig. 8. Experimental setup of the reciprocating compressor: (a) data acquisition system, (1) exhaust valve, (2) accelerometer on the exhaust valve lid, (3) data acquisition box, (4) control panel, (b) diagram of reciprocating compressor, and (c) two reciprocating compressor faults of valve wear and valve fracture.

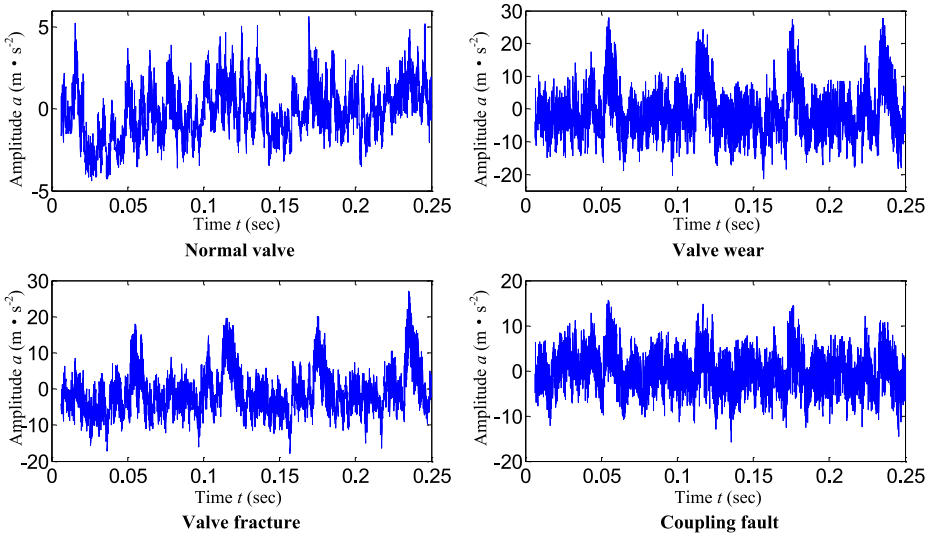

Fig. 9. The waveforms of four data sets including normal valve, valve wear, valve fracture and the coupling fault of them.

The time-domain waveforms of them are depicted in Fig. 9. It is clear that there exists strong background noise in field data. There are 1000 data samples of length 4000 in each subset come from the whole collected data. 50 samples are used for training and the rest 50 samples are used for prediction.

5. Results and discussions

For classification tasks two classifiers were employed in this research work: (a) the nearest neighbor (NN) classifier, which is a simple parameter-free classifier and (b) the support vector machine (SVM) classifier, which is a classical parameterized classifier. We chose these two relatively simple classifiers, because in this study we basically focus on the application of ISAX approach for feature extraction. The advantages of the features generated by the ISAX method over the traditional features are mainly investigated rather than the selection of classifier.

Generally, parameters of KPCA are set by trial and error. In this paper, the Gaussian RBF kernel with $\sigma = 100$ was found to be the optimal kernel function according to the model providing the maximum cluster separation and minimum cohesion [38]. The reduced dimension was selected based on the accumulated contribute rate of feature variances. The threshold was set to be 95% in this paper.

It is clear that the performance of SVM heavily depends on the choices of parameters. In this study, 10-fold cross validation and Gaussian kernel function were selected to evaluate the classification performance. The optimal values of parameters were selected using grid search method. The optimal value for cost parameter c was selected for the range of $\left\{2^i | i = -4, -2, 0, 1, 2, 4, 6, 8\right\}$ and the Gaussian kernel parameter γ was selected for the range of

$\left\{2^i | i = -4, -2, 0, 2, 4, 6, 8\right\}$. The parameter combination of c and γ with best cross-validation accuracy was picked.

5.1. Diagnostic performance with experimental data

For comparison reasons, we firstly performed experiments using the 13 features generated by the traditional methods mentioned in Section 4.1.2 with two classifiers. These features were directly used as the inputs to the classifiers for fault diagnosis. The classification results in Fig. 10, reveal that the classification tasks using common features perform poorly with a relatively low classification rate. It also can be seen that the performance using SVM classifier is superior to that using NN classifier because of the parameter optimization scheme in SVM classifier.

Next, the experiments using the ISAX method were carried out to verify its superiorities. The 13 common features were used as the coordinates in PAA process of ISAX method and the principal features generated by KPCA were used as inputs of the SVM and NN classifiers. Four classification datasets described in Table 2 were carried out and the results are summarized in Tables 4–7 in the form of confusion matrices corresponding to the classification problem A, B, C and D, respectively. Take Tables 4 and 5 for a detailed description, the row heading represents the estimated class and the column heading represents the true class. There are 4 types of defects for the problems A and B. In Table 4, all the testing samples are classified correctly and the accuracy rate is 100%. In Table 5, 100 testing samples of the defect of 2B are used for classification with SVM, in which 92 samples are classified as the defect of 2B correctly. 1 sample is misclassified as the defect of 1B and 2 samples are misclassified as the defect of 3B and 5 samples are misclassified as the defect of 4B. So the accuracy rate of the defect of 2B is 92%. Similarly, the classification rates of defect of 3B and 4B are 92% and 97.33%. And the classification rate of the problem of B is the average of the rates of the 3 types of subproblems.

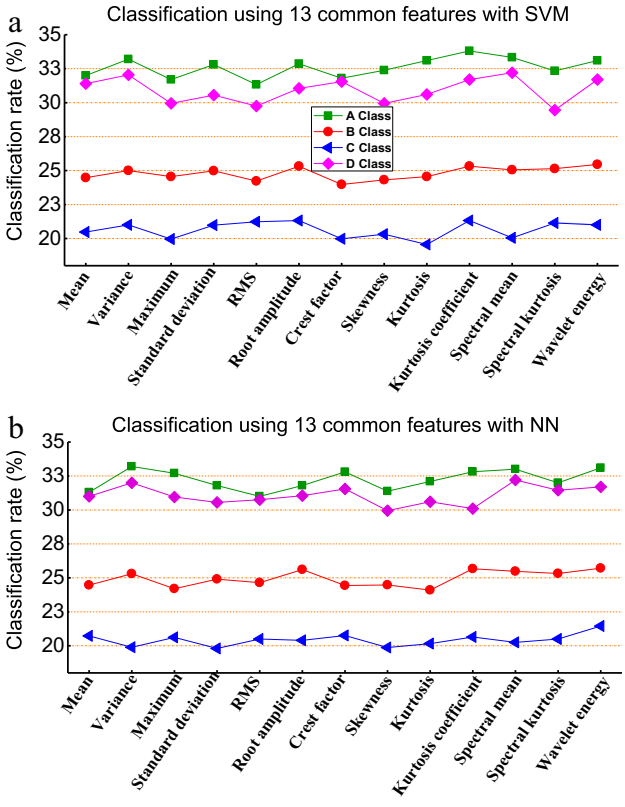

Fig. 10. Classification rates for A-D classification problems with SVM and NN classifiers using data from Case Western Reserve University Bearing Data Center.

Table 4
Confusion matrix for the classification problem A for the two classifiers of SVM and NN.

SVM/NN		Estimated class			
		1A	2A	3A	4A
True class	1A	400/400	0/0	0/0	0/0
	2A	0/0	100/100	0/0	0/0
	3A	0/0	0/0	150/150	0/0
	4A	0/0	0/0	0/0	300/300

Table 5
Confusion matrix for the classification problem B for the two classifiers of SVM and NN.

SVM/NN		Estimated class			
		1B	2B	3B	4B
True class	1B	–/–	–/–	–/–	–/–
	2B	1/1	92/95	2/2	5/2
	3B	0/0	2/8	145/140	3/2
	4B	0/0	2/11	6/7	292/282

Table 6
Confusion matrix for the classification problem C for the two classifiers of SVM and NN.

SVM/NN		Estimated class			
		1C	2C	3C	4C
True class	1C	–/–	–/–	–/–	–/–
	2C	0/0	98/77	2/20	0/3
	3C	0/0	22/20	110/101	19/30
	4C	0/0	6/30	57/67	237/203

According to Fig. 10 and Tables 4–7, the classification rates of 4 problems using traditional methods and ISAX are summarized in Fig. 11. The overall classification rate using traditional methods is the average of the ones using 13 features. The overall classification rate of every problem is also the average of the ones of subproblems. To analyze the performance of the original SAX method, the experiments using SAX method mentioned in [30] following the above procedure were carried out and the results were also given in Fig. 11. The features generated by SAX are the inputs of classifiers.

From Fig. 11, specific conclusions can be seen. For the problem of A, the diagnostic scheme based on ISAX performs so well as to reach up to 100% accuracy with both classifiers because of the simple classification problems. The classification rate of ISAX method with SVM classifier (NN classifier) is 67.40% (67.77%) higher than traditional methods and is 12.75% (21.25%) higher than the SAX method. For the problem of B, the classification rate of ISAX method with SVM classifier (NN classifier) is 69.57% (69.16%) higher than traditional methods and is 18.37% (19.86%) higher than the SAX method. The results indicate that the ISAX method trained by the incipient fault data can diagnose serious faults better. For the problem of C, the classification rate of ISAX method with SVM classifier (NN classifier) is 60.15% (50.24%) higher than traditional methods and is 30.60% (18.24%) higher than the SAX method. It is suggested that the ISAX method trained by serious fault data can diagnose incipient faults with a better performance compared with other methods. For the problem of D, though it is a relatively difficult classification task, the classification rate of ISAX method with SVM classifier (NN classifier) is 68.49% (68.13%) higher than traditional methods and is 19.25% (22.04%) higher than

Table 7
Confusion matrix for D classification task with two classifiers.

SVM/NN		Estimated class									
		1D	2D	3D	4D	5D	6D	7D	8D	9D	10D
True class	1D	400/400	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
	2D	0/0	99/99	0/0	0/0	1/0	0/0	0/0	0/1	0/0	0/0
	3D	0/0	0/0	148/149	0/1	0/0	2/0	0/0	0/0	0/0	0/0
	4D	0/0	0/0	0/0	300/299	0/0	0/0	0/1	0/0	0/0	0/0
	5D	0/0	0/0	0/0	0/1	97/96	2/1	0/0	0/1	0/0	1/1
	6D	0/0	0/0	0/0	0/0	0/0	150/150	0/0	0/0	0/0	0/0
	7D	0/0	0/0	0/0	0/0	0/0	0/0	300/300	0/0	0/0	0/0
	8D	0/0	0/0	0/0	0/0	1/0	0/0	0/0	98/98	1/1	0/1
	9D	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	150/150	0/0
	10D	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	300/300

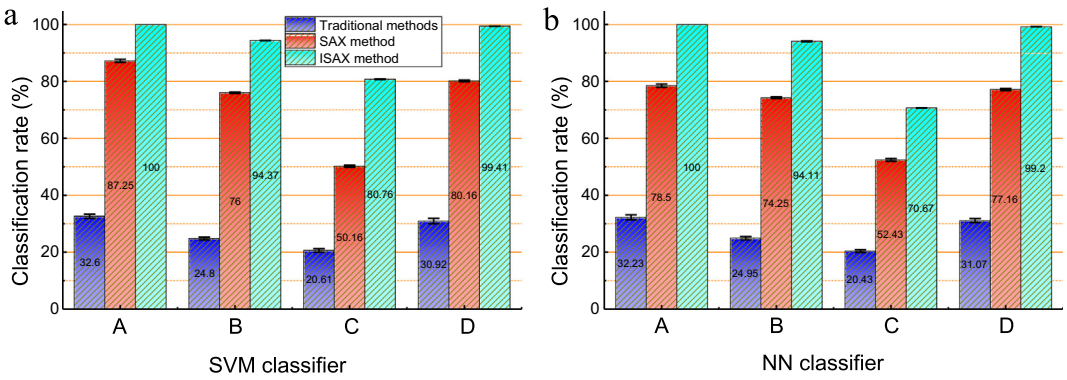

Fig. 11. Summary comparison results as well as standard error bars using traditional method, SAX method and ISAX method with different classifiers.

Table 8
Confusion matrix for four classification tasks using the RMS and the ISAX based features.

SVM		Estimated class			
		N	V_w	S_f	C_f
True class	N	326/402	45/27	101/45	28/26
	V_w	58/12	241/435	92/40	109/13
	S_f	79/9	89/7	223/478	109/6
	V_f	56/8	147/30	60/6	237/456

the SAX method. This implies that the ISAX method can identify different defect categories as well as the different severities better. For an overall comparison, the classification rate of ISAX method (the average accuracy for problems of A-D) with SVM classifier (NN classifier) is 66.40% (63.83%) higher than traditional methods and is 20.24% (20.41%) higher than the SAX method.

5.2. Diagnostic performance with field data

In this section, the common feature of RMS was used for classification tasks for the comparison experiments. Since the SVM classifier seems to outperform the NN, only the SVM classifier was conducted.

The experiment results using field data were depicted in Table 8 with confusion matrix. To show the experimental results more clearly, Fig. 12 was given to imply the classification and misclassification rate. The ordinate is the true classification and the abscissa is the estimated classification. So the diagonal data represents the classification rate, and others represent the misclassification rate. From Fig. 12, it can be seen that the classification rate using RMS is in the range of 44.60%–65.20% (average 54.90%) and the one using ISAX based features is 80.40%–95.60% (average 88.00%). It is clear that the presented scheme achieves impressive performance with a remarkable classification accuracy in spite of strong background

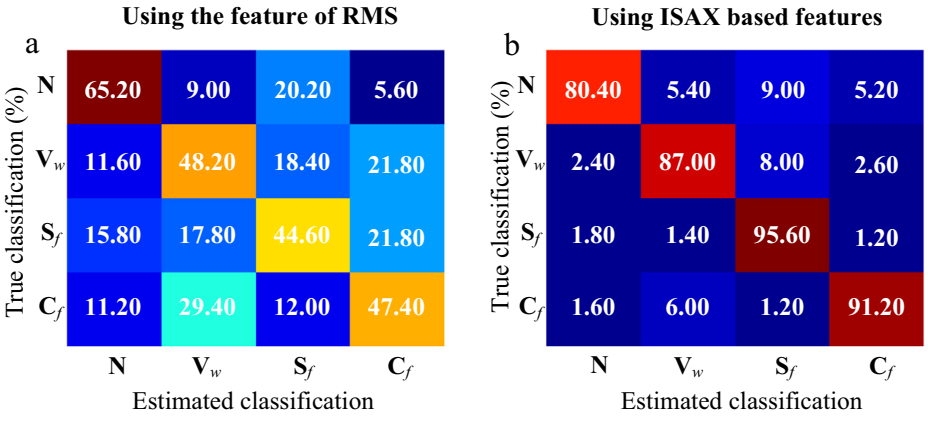

Fig. 12. Overall accuracy of the four classification problems with SVM classification.

noise. For further analysis, the symbolic approximation representation has a fairly good ability of noise reduction and meanwhile retains fault information as much as possible. The ISAX method investigates signal characteristics from multi-domain for a comprehensive feature extraction and achieves better performance than the traditional methods. In addition, the data compression is achieved along with the symbol representation process, and it has an obvious advantage over the existing feature extracted methods especially in online real-time detection and diagnosis.

All the above result analyses prove that the ISAX method has obtained significant achievements in recognition accuracy and provided a better generalisation capability in comparison with the commonly used feature extraction methods. It is more effective to apply the ISAX method into feature extraction for machinery fault diagnosis.

6. Conclusions

An intelligent method of feature extraction using ISAX has been proposed for condition monitoring and fault diagnosis. The novel contribution to this study is to improve the SAX algorithm for extracting comprehensive signal characteristics from multi-domain. Moreover, local structure analysis is conducted on the symbol sequences to present the fault patterns. Experimental studies on faulty rolling bearings and reciprocating compressor valves demonstrated the validity of the proposed method. The conclusions can be drawn as follows.

(1) To extract comprehensive fault information, the SAX is improved to create various symbol sequences based on multi-domain features. The ISAX based features led to an impressive performance with a remarkable classification accuracy.
(2) The fault patterns are firstly proposed to investigate the sequential varying characteristics among neighbored data points. The experimental studies demonstrated its superiorities compared with the traditional feature extraction methods.
(3) By integrating the ISAX, Shannon entropy and KPCA, a new framework of machinery fault diagnosis was investigated and validated.

Future scope of work can be directed toward to the fault patterns related to different typical defects and the robustness of the presented method under strong background noise.

Acknowledgements

This research acknowledges the financial support provided by National Key R&D Program of China, China (Grant No. 2016YFC0303700) and National Natural Science Foundation of China, China (Grant No. 51674277).

Appendix A. The principle of KPCA

The basic idea of this algorithm is to map the raw data into a high-dimensional feature space via defining a nonlinear function ϕ and then employ linear PCA analysis in space F.

Let $x_1, x_2, x_3, \cdots, x_n \in R^M$. Though the transformation by using the nonlinear function ϕ, the data x_i is extended as $\phi(x_i)$ in high-dimensional feature space: $x_i \in R^M \rightarrow \phi(x_i) \in R^N$ The correlation matrix of the feature space can be calculated by

$$C = \frac{1}{n}\sum_{i=1}^{n}\phi(x_i)\phi(x_i)^T \tag{A.1}$$

The eigenvalue λ and eigenvector μ of the covariance matrix C are given by

$$\lambda\mu = C\mu \tag{A.2}$$

Then λ and μ can be expanded by

$$\lambda a = \frac{K}{N}a \tag{A.3}$$

where $a = [a_1, a_2, a_3, \cdots, a_n]$ and $K_{ij} = \phi(x_i)^T\phi(x_j) = \phi(x_i)\phi(x_j)$ is a kernel matrix in space F. Regarding the kernel functions, the radial basis function (RBF) kernel gives a good performance when the optimal parameter is used, which is given by

$$K_{ij} = \frac{\exp(-||x_i - x_j||)^2}{2\sigma^2} \tag{A.4}$$

where $2\sigma^2 = w$ is the width of the Gaussian kernel.

The most representative d dimensional features extracted by KPCA form the following vector:

$$\mathrm{X} = \left[\frac{\sum_{j=1}^{N}a_j^{(1)}k(x_j, x_{new})}{\sqrt{\lambda_1^a}}, \frac{\sum_{j=1}^{N}a_j^{(2)}k(x_j, x_{new})}{\sqrt{\lambda_2^a}}, \ldots, \frac{\sum_{j=1}^{N}a_j^{(d)}k(x_j, x_{new})}{\sqrt{\lambda_d^a}}\right] \tag{A.5}$$

where $a^{(1)}, a^{(2)}, \cdots, a^{(d)}$ are the d eigenvectors corresponding to the first largest d eigenvalues $\lambda_1^a, \lambda_2^a, \cdots, \lambda_d^a$ according to the Eq. (A.2). $a_j^{(i)}$ is the j-th component of the vector $a^{(i)}$.

References

[1] R. Teti, K. Jemielniak, G. O'Donnell, D. Dornfeld, Advanced monitoring of machining operations, CIRP Ann. – Manuf. Technol. 59 (2) (2010) 717–739.
[2] R.U. Maheswari, R. Umamaheswari, Trends in non-stationary signal processing techniques applied to vibration analysis of wind turbine drive train - A contemporary survey, Mech. Syst. Sig. Process. 85 (2017) 296–311.
[3] B. Christian, A. Gläser, The behavior of different transformer oils relating to the generation of fault gases after electrical flashovers, Int. J. Electr. Power Energy Syst. 84 (2017) 261–266.
[4] Y. Wang, C. Xue, X. Jia, X. Peng, Fault diagnosis of reciprocating compressor valve with the method integrating acoustic emission signal and simulated valve motion, Mech. Syst. Sig. Process. 56 (2015) 197–212.
[5] Z. Du, X. Jin, Y. Yang, Fault diagnosis for temperature, flow rate and pressure sensors in VAV systems using wavelet neural network, Appl. Energy 86 (9) (2009) 1624–1631.
[6] B. Subramainam, B.B. Lahiri, T. Saravanan, J. Philip, T. Jayakumar, Infrared thermography for condition monitoring-a review, Infrared Phys. Technol. 60 (2013) 35–55.
[7] A. Benammar, R. Drai, A. Guessoum, Ultrasonic flaw detection using threshold modified S-transform, Ultrasonics 54 (2) (2014) 676–683.
[8] M. Amar, I. Gondal, C. Wilson, Vibration spectrum imaging: A novel bearing fault classification approach, IEEE Trans. Ind. Electron. 62 (1) (2015) 494–502.
[9] H. Liu, J. Zhang, Y. Cheng, C. Lu, Fault diagnosis of gearbox using empirical mode decomposition and multi-fractal detrended cross-correlation analysis, J. Sound Vib. 385 (2016) 350–371.
[10] A.B. Ming, Z.Y. Qin, W. Zhang, F.L. Chu, Spectrum auto-correlation analysis and its application to fault diagnosis of rolling element bearings, Mech. Syst. Sig. Process. 41 (1) (2013) 141–154.
[11] Y. Maouche, M.E.K. Oumaamar, M. Boucherma, A. Khezzar, Instantaneous power spectrum analysis for broken bar fault detection in inverter-fed six-phase squirrel cage induction motor, Int. J. Electr. Power Energy Syst. 62 (2014) 110–117.
[12] W.B. Wu, S.Q. Yang, Y.J. Huang, Fault diagnosis of pressure relief valve based on power spectrum, Energy Procedia 17 (2012) 1319–1325.
[13] B. Liang, S.D. Iwnicki, Y. Zhao, Application of power spectrum, cepstrum, higher order spectrum and neural network analyses for induction motor fault diagnosis, Mech. Syst. Sig. Process. 39 (1) (2013) 342–360.
[14] R.L. Rosero, L. Romeral, J.A. Ortega, E. Rosero, Short-circuit detection by means of empirical mode decomposition and Wigner-Ville distribution for PMSM running under dynamic condition, IEEE Trans. Ind. Electron. 56 (11) (2009) 4534–4547.
[15] Z.K. Peng, F.L. Chu, Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography, Mech. Syst. Sig. Process. 18 (2) (2004) 199–221.
[16] Y.G. Lei, J. Lin, Z.J. He, M.J. Zuo, A review on empirical mode decomposition in fault diagnosis of rotating machinery, Mech. Syst. Sig. Process. 35 (1) (2013) 108–126.

[17] R.J. Alfredson, J. Mathew, Time domain methods for monitoring the condition of rolling element bearings, NASA STI/Recon Technical Report A 86 (1985) 22166.
[18] A.K.S. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mech. Syst. Sig. Process. 20 (7) (2006) 1483–1510.
[19] J. Zheng, H. Pan, S. Yang, J. Cheng, Adaptive parameterless empirical wavelet transform based time-frequency analysis method and its application to rotor rubbing fault diagnosis, Signal Process. 130 (2017) 305–314.
[20] N.E. Hang, The empirical mode decomposition method and the hilbert spectrum for non-stationary time series, Proc. Roy. Soc. London 45AA (1998) 703–775.
[21] Y. Huang, D. Wu, Z. Zhang, H. Chen, S. Chen, EMD-based pulsed TIG welding process porosity defect detection and defect diagnosis using GA-SVM, J. Mater. Process. Technol. 239 (2017) 92–102.
[22] W.Y. Duan, Y. Han, L.M. Huang, B.B. Zhao, M.H. Wang, A hybrid EMD-SVR model for the short-term prediction of significant wave height, Ocean Eng. 124 (2016) 54–73.
[23] V. Bono, S. Das, W. Jamal, K. Maharatna, Hybrid wavelet and EMD/ICA approach for artifact suppression in pervasive EEG, J. Neurosci. Methods 267 (2016) 89–107.
[24] J. Gilles, Empirical wavelet transform, IEEE Trans. Signal Process. 61 (16) (2013) 3999–4010.
[25] H. Cao, F. Fan, K. Zhou, Z. He, Wheel-bearing fault diagnosis of trains using empirical wavelet transform, Measurement 82 (2016) 439–449.
[26] M. Kedadouche, Z. Liu, V.H. Vu, A new approach based on OMA-empirical wavelet transforms for bearing fault diagnosis, Measurement 90 (2016) 292–308.
[27] J. Pan, J. Chen, Y. Zi, Y. Li, Z. He, Mono-component feature extraction for mechanical fault diagnosis using modified empirical wavelet transform via data-driven adaptive Fourier spectrum segment, Mech. Syst. Sig. Process. 72 (2016) 160–183.
[28] J. Chen, J. Pan, Z. Li, Y. Zi, X. Chen, Generator bearing fault diagnosis for wind turbine via empirical wavelet transform using measured vibration signals, Renew. Energy 89 (2016) 80–92.
[29] A.V. Oppenheim, R.W. Schafer, J.R. Buck, Discrete-time Signal Processing, Prentice Hall, 1999.
[30] G. Georgoulas, P. Karvelis, T. Loutas, C.D. Stylios, Rolling element bearings diagnostics using the Symbolic Aggregate approXimation, Mech. Syst. Sig. Process. 60 (2015) 229–242.
[31] J. Lin, E. Keogh, S. Lonardi, B. Chiu, A symbolic representation of time series, with implications for streaming algorithms, in: Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, San Diego, CA, 2003.
[32] B. Lkhagva, Y. Suzuki, K. Kawagoe, Extended SAX: Extension of symbolic aggregate approximation for financial time series data representation. DEWS2006 4A-i8, 7 (2006).
[33] SAX Homepage, <http://www.cs.ucr.edu/~eamonn/SAX.htm> (last accessed 15.03.2017).
[34] J. Shieh, E. Keogh, i SAX: indexing and mining terabyte sized time series, in: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2008, pp. 623–631.
[35] C.E. Shannon, A mathematical theory of communication, ACM SIGMOBILE Mobile Comput. Commun. Rev. 5 (1) (2001) 3–55.
[36] B. Schölkopf, A. Smola, K.R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (5) (1998) 1299–1319.
[37] K.A. Loparo, Bearing vibration dataset, Case Western Reserve University, 2014 <http://csegroups.case.edu/bearingdatacenter/pages/welcome-casewestern-reserve-university-bearing-data-center-website> (last accessed 17.03.2017).
[38] M.B. Richman, I. Adrianto, Classification and regionalization through kernel principal component analysis, Phys. Chem. Earth. 35 (9–12) (2010) 316–328.