

Research Article

Association Rule-Based Feature Mining for Automated Fault Diagnosis of Rolling Bearing

Yuan Li ^{1,2}, Jinjiang Wang ³, Lixiang Duan ³, Tangbo Bai,³ Xuduo Wang,³ Yulong Zhang,³ and Guoliang Qin¹

¹School of Energy and Power Engineering, Xi'an Jiaotong University, Xi'an 710049, China

²Sinopec Si Chuan to Eastern China Transmission Gas Pipeline Co., Ltd., Wuhan 430020, China

³School of Safety and Ocean Engineering, China University of Petroleum, Beijing 102249, China

Correspondence should be addressed to Jinjiang Wang; jwang@cup.edu.cn

Received 13 February 2019; Revised 10 June 2019; Accepted 18 July 2019; Published 25 August 2019

Academic Editor: F. Viadero

Copyright © 2019 Yuan Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Effective and efficient diagnosis methods are highly demanded to improve system reliability. Comparing with conventional fault diagnosis methods taking a forward approach (e.g., feature extraction, feature selection, and fusion, and then fault diagnosis), this paper presents a new association rule mining method which provides an inverse approach unearthing the underlying relation between labeled defects and extracted features for bearing fault analysis. Instead of evenly dividing methods used in traditional association rule mining, a new association rule mining approach based on the equal probability discretization method is presented in this study. First, a series of extracted features of signal data are discretized following the guideline of equalized probability distribution of the data in order to avoid excessive concentration or decentralized data. Next, the data matrix composed of arrays of discretized features and defect labels is exploited to generate the association rules representing the relation between the features and fault types. Experimental study on a bearing test reveals that the proposed method can generate a series of underlying association rules for bearing fault diagnosis, and the related features selected by the proposed method can be used directly to analyze bearing signals for fault classification and defect severity identification. As a new feature selection method, it possesses prominent superiority compared to traditional PCA, KPCA, and LLE dimension reduction methods.

1. Introduction

The machinery reliability is critical to the system operational safety. Rolling element bearing, as the widely used component in large mechanical system, plays an important role in ensuring the availability of machineries such as aircraft engine, wind turbine, and compressor [1]. Due to the harsh operating conditions (e.g., high speed, heavy load, and great heat), bearings may lead to a sudden and catastrophic failure [2]. If it fails to diagnose earlier, bearing fault may incur great losses, and even a terrible accident. Effective algorithms for bearing defect diagnosis and prognosis are demanded and have remained an active research field to increase system reliability [3].

Fault diagnosis is the problem of detecting the potential faults hidden in the observed instances that are related to

specific application domains [4]. Extensive efforts have been made by taking a forward approach for bearing fault diagnosis including feature extraction, feature selection, and fusion, and then fault diagnosis modeling. Various signal processing techniques including wavelet transform [5–7], empirical mode decomposition [8], order tracking [9], and spectral analysis [10, 11] have been investigated for incipient defect feature extraction and diagnosis. An integrative algorithm of sparse coding and online dictionary learning is developed in [2] to extract impulse features for machinery fault detection. Then, the extracted features are selected or fused to build a data-driven model based on artificial intelligence techniques including neural network [12, 13], support vector machine [14–16], and fuzzy c-means [17, 18] for bearing fault classification. In such a forward approach, feature extraction plays a key role in the model for bearing

defect diagnosis. The feature extraction is usually performed empirically based on prior experience. Thus, it lacks a systematic manner for bearing defect signature analysis.

To address the above issues, the association rule mining method provides a new approach for bearing defect signature analysis [19]. It takes an inverse approach to search the relevance and association of large information and has been investigated in commerce [20, 21], traffic [22], tourism [23], biomedical applications [24, 25], power plant equipment diagnosis [26], and the analysis of telecommunication networks [27]. In the association rule mining technique, data discretization is a critical step to find the quantitative attributes from the relation tables of potential items. In terms of data discretization, the equal density discretization and equal width discretization are commonly used uniform partitioning approaches. However, such approaches neglect the probability distribution characteristics of these features. Thus, it may make data unbalance, such as excessive concentration or decentralized data, and generate unsatisfactory association rules [28].

In line with the above challenges, this paper presents a new association rule mining method based on equal probability for bearing defect features analysis. First, a series of extracted features of signal data are discretized following the guideline of equalized probability distribution of the data in order to avoid excessive concentration or decentralized data. To evaluate the data discretization performance, the new criteria named information entropy of interval class is formulated. Next, the data matrix composed of arrays of discretized features and defect labels is exploited to generate the association rules representing the relation between features and fault types. The generated rules are then used for bearing defect classification based on fuzzy proximity methods and feature selection that generating the representative features related to typical defects. The related features selected by the proposed method can be used directly to analyze signals for fault classification and defect severity identification avoiding the impact of irrelevant features on the premise of keeping the original feature state. An experimental study is performed to validate the effectiveness of the presented method using the bearing test data provided by Case Western Reserve University (CWRU), and the experimental results reveal that the new method can effectively generate a series of underlying association rules for bearing fault diagnosis, and yields the best discretization performance and classification accuracy. The related features selected by the proposed method can be used directly to analyze bearing signals for fault classification and defect severity identification avoiding the impact of irrelevant features on the premise of keeping the original feature state. The proposed method outperforms the traditional principal component analysis (PCA) [29], kernel principal component analysis (KPCA) [30], and locally linear embedding (LLE) [31] dimension reduction methods [32, 33].

The intellectual merits of this paper rest on two folds. (1) A new association rule mining method with the data discretization of equal probability distribution is firstly presented, and a new criterion of information entropy of interval class is also formulated. (2) The presented method,

as an inverse approach for bearing defect signature analysis, provides a new tool to guide the feature extraction instead of empirical feature extraction in the current forward approach of bearing defect diagnosis. The rest of the paper is structured as follows. In Section 2, the theoretical framework is introduced including the association rule mining technique and the Apriori algorithm. The proposed method is discussed in detail in Section 3. The information entropy criteria of interval class are also formulated to assess the performance of data discretization. In Section 4, the effectiveness of the presented method is demonstrated using bearing test data provided by CWRU. Finally, the conclusions are drawn in Section 5.

2. Related Work

Many studies published in the literature adopt association rule mining to find useful knowledge from database proactively. The discovered knowledge with association rules can be applied to information management, decision making, process control, and many other applications.

2.1. Association Rule Mining. Association rule mining is a technique to detect and extract meaningful association relationships hidden in databases. It is firstly introduced by Agrawal et al. [34] and has been investigated in different applications including commodity sales [35], disease study [36], quality improvement of a production process [37], and alarm correlation analysis [38]. The association rule mining method is formulated as follows.

Let $I = \{i_1, i_2, \dots, i_m\}$ be a set of literals, referred to as items. Let $D = \{t_1, t_2, \dots, t_n\}$ be a set of transactions. Each transaction t in D has a unique transaction TID and contains a subset of items I' where $I' \subseteq I$. An association rule is defined as an implication of the form $X \implies Y$ where $X, Y \subset I$, and $X \cap Y = \emptyset$. The support for an itemset X ($\text{supp}(X)$) is defined as the proportion of transactions in the transactions which contains the itemset. Itemsets with the minimum support are called large itemsets and all others small itemsets. The confidence of a rule is defined as $\text{conf}(X \implies Y) = \text{supp}(X \cup Y) / \text{supp}(X)$. Therefore, the association rule $X \implies Y$ will satisfy:

$$\begin{aligned} \text{supp}(X \cup Y) &\geq \sigma, \\ \text{conf}(X \implies Y) &\geq \delta, \end{aligned} \quad (1)$$

where σ and δ are the minimum support and confidence, respectively.

Association rules are typically summarized as two steps. First, find all large itemsets that have transaction support above the minimum support. Then each large itemset is used to generate the desired rules, which satisfies the minimum confidence constraint.

2.2. Apriori Algorithm. As a classical algorithm, Apriori algorithm discovers the frequent itemsets which make enormous passes over the data. In the first pass, the frequent itemsets are calculated by finding the support of individual

items with the minimum support. In each subsequent pass, a seed set of itemsets which are found to be large in the previous pass are taken as the objects. This seed set is used to generate new potentially large itemsets, called candidate itemsets. The actual support is counted for these candidate itemsets during the passes over the data. At the end of the pass, the actual large candidate itemsets become the seed for the next pass. This process continues until no new frequent itemsets are found. The Apriori algorithm generates the candidate itemsets without considering the transactions in the database to improve the computational efficiency.

3. The Proposed Method

In bearing defect diagnosis, it is a challenge to analyze the bearing defect signatures in a systematic manner. In order to recognize bearing fault efficiently and accurately, this paper presents a new equal probability-based association rule mining method for the bearing defect signature analysis method.

3.1. Formulation of the Proposed Approach. The framework of the presented method consists of four different modules including data acquisition, feature extraction, equal probability-based association rule mining, association analysis, and fault diagnosis as shown in Figure 1. First, it collects normal and fault signal data from the monitoring equipment and extract features in time and frequency domains from the obtained data. Then, the extracted features are discretized and transformed into symbolic sequences. Next, the relation between discretized features and defect modes labeled is used to formulate the rules. Finally, the representative features related to typical defects are extracted and investigated based on the rules, which can not only be used to classify the bearing in different conditions, but also provide a guiding significance to traditional bearing fault diagnosis.

As a new feature selection method, it is different from other data dimensionality reduction methods, such as PCA, KPCA, and LLE, which need to transform the data matrix leading to bereave of the actual physical meaning of the selected features. The related features selected by the proposed method can be used directly to analyze the fault to avoid the impact of irrelevant features on the premise of keeping the original feature state. The range of eigenvalues obtained from the representative features can be used to determine the type and size of fault according to the respective values of features. This makes it possible to assess bearing status directly from sensing measurements instead of relying on complex models in conventional bearing defect diagnosis.

3.2. Association Rule Mining Based on Equal Probability. Typically, it is required that the data follow Boolean attributes, such as “0” and “1,” for association rule mining [39]. However, the relational tables in most business and scientific domains have the rich attribute forms such as quantitative attributes (e.g., age and income), while the Boolean attributes can be considered as a special case of quantitative attributes [39]. To solve such a problem, a simple approach is to partition the values into intervals and then map each

interval into a Boolean attribute. Therefore, data discretization becomes an essential step for mining association rules to transform data from quantitative attributes to Boolean attribute. While inspired by symbolic aggregate approximation (SAX, a symbolic representation of sequential data) [40], this paper proposes a new association rule mining method based on equal probability distribution. According to the characteristics of data distribution, this method divides the sequence into several intervals according to the criterion of equal probability distribution. First, the sequence is standardized and the normalized sequence is subject to the Gaussian distribution, $X \sim N(0, 1)$. The equation is as follows:

$$B = \frac{A - \mu}{\sigma}, \quad (3)$$

where B is the normalized array of A , μ is the mean of the sequence A , and σ is the standard deviation.

When the data obey the Gaussian distribution, the probability that the data points fall within the range of $[a, b]$ is the area surrounded by the standard Gaussian distribution curve, as shown in Figure 2. The probability formula is described as

$$P(a \leq X \leq b) = \phi(x) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)} dx. \quad (4)$$

According to the characteristics of the Gaussian distribution, the data can be graded in the form of the equal probability distribution. Then, the data matrix consists of an array of features in different levels and a series of labels that represent different states.

Next, the association relation between the discretized features and labeled defect modes is drilled to formulate the rules. Finally, the sensitive features related to typical defects are extracted and investigated according to the rules. The presented method supports the bearing status assessment directly from sensing measurements instead of relying on complex models in the traditional fault diagnosis approach.

3.3. Feature Extraction and Discretization Effect Assessment Method. A total of 12 commonly used bearing vibration signal features are extracted, including eight features from time domain and four features from frequency domain, as listed in Table 1. These features are often used to depict the waveform, mutation, and distribution characteristics of the bearing vibration signal for bearing fault diagnosis. In order to compare the discretization performance of different methods and demonstrate the effectiveness of the discretization represented in this paper, the interval class-information entropy criteria is introduced as

$$\text{Ent}(I_i) = \sum_{j=1}^k \left[\left(\frac{c_{ij}}{r_i} \right) * \log_2 \left(\frac{c_{ij}}{r_i} \right) \right]. \quad (5)$$

Typically, the interval class-information entropy reflects the category diversity of one interval. Normally, the larger the interval class-information entropy, the worse the discretization performance.

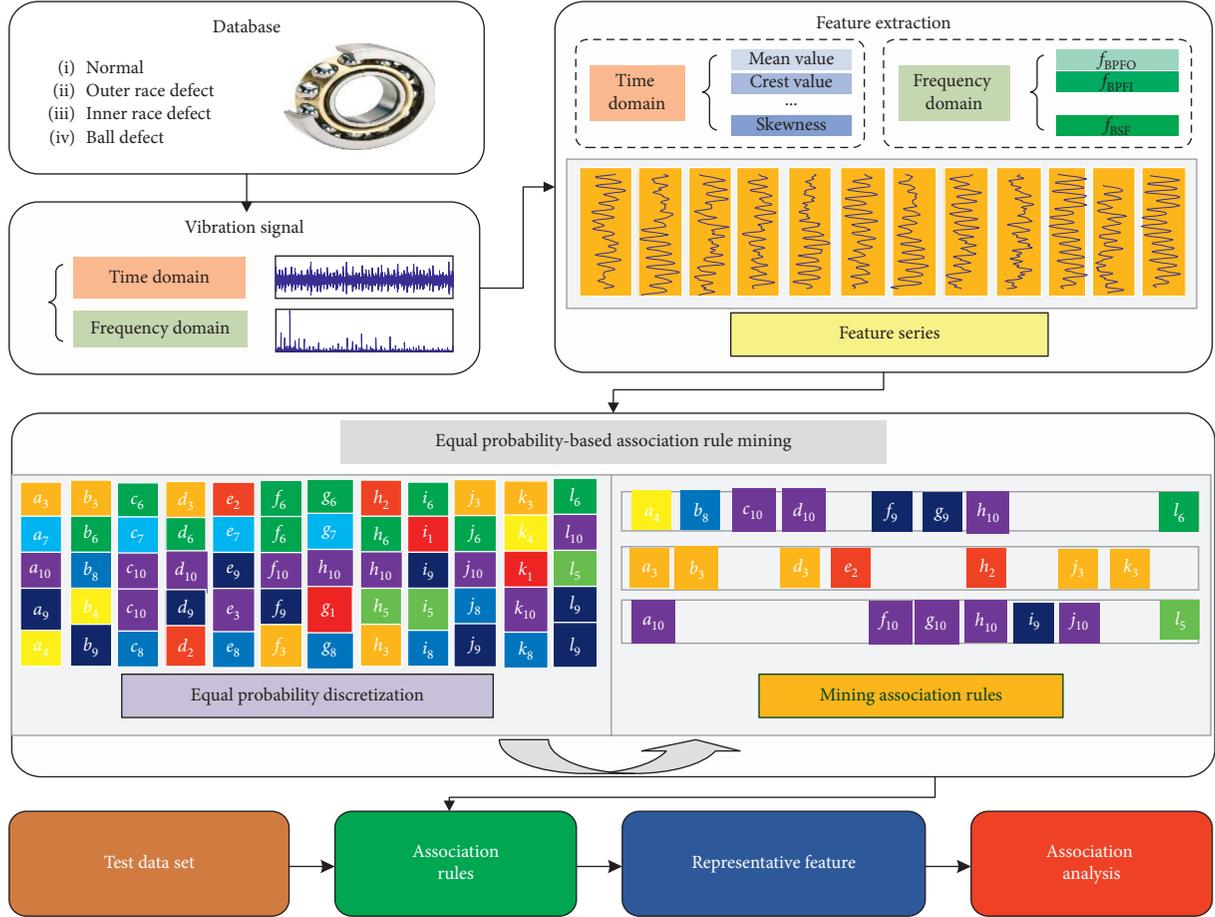


FIGURE 1: The framework of association rule mining for bearing defect diagnosis.

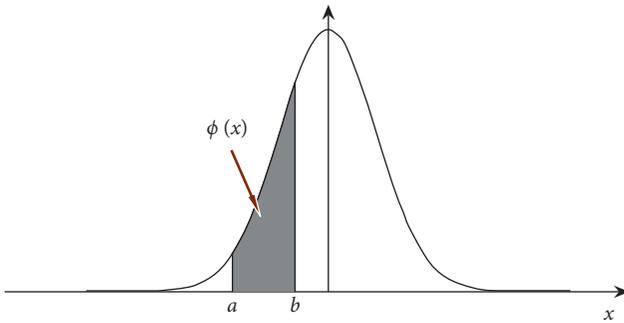


FIGURE 2: The probability of Gaussian distribution.

In order to evaluate the performance of the proposed interval class-information entropy criteria, the common RMS-based indicator is introduced to compare the discretization performance of different methods. The indicator is described as follows:

$$BRI_i = \frac{r_i}{\sqrt{(\sum_{j=1}^k c_{ij}^2)/k}} \quad (6)$$

where BRI_i is the RMS-based discretization performance indicator of evaluating different methods. A larger calculation

result leads to more different kinds of data that exist in an interval, and the discretization method shows a worse performance.

4. Experimental Studies

4.1. Experimental Setup and Dataset. The experimental data were provided by Case Western Reserve University [41], and the experimental setup is shown in Figure 3. A motor drives a shaft via a dynamometer and electronic control system. The test data used in this study come from deep groove ball bearings (6205-2RS JEM SKF) installed in the motor-driven mechanical system at the drive end of the motor. The failures of the test bearings were set using an electrodischarge machining (EDM) with single point faults. For each test, vibration data were collected through accelerometers attached to the housing with a magnet at the drive end with a sampling rate of 12,000 Hz. The motor speed is 1,797 r/min, and the theoretical shaft frequency is 29.95 Hz. There are three types of fault datasets for outer and inner race faults and rolling element faults, and each fault is grouped into three categories according to the fault diameters of 0.007, 0.014, and 0.021 inches. Considering the normal state, there are 10 types of data samples in total.

TABLE 1: List of extracted features.

Domain	Features	Expression
Time	Mean value	$x_{MV} = (1/n)\sum_{i=1}^n x_i$
	Standard deviation	$x_{SD} = \sqrt{(\sum_{i=1}^n (x_i - x_{MV})^2)/(n-1)}$
	Root mean square	$x_{RMS} = \sqrt{(\sum_{i=1}^n x_i^2)/n}$
	Root amplitude	$x_{RA} = (\sqrt{(\sum_{i=1}^n x_i^2)/n})^2$
	Crest value	$x_{CF} = \max_{1 \leq i \leq n} x_i$
	Skewness	$x_s = (\sum_{i=1}^n x_i^3)/n$
	Kurtosis	$x_K = (\sum_{i=1}^n x_i^4)/n$
Frequency	Kurtosis coefficient	$x_{KC} = x_K/x_{RMS}^4$
	Ball pass frequency, inner race	$f_{BPFI} = (Nf_r/(60 \times 2))(1 - (d/D)\cos \phi)$
	Ball pass frequency, outer race	$f_{BPFO} = (Nf_r/(60 \times 2))(1 + (d/D)\cos \phi)$
	Ball (roller) spin frequency	$f_{BSF} = (Df_r/(60 \times 2d))(1 - ((d/D)\cos \phi)^2)$
	Fundamental frequency	$f = f_r/60$

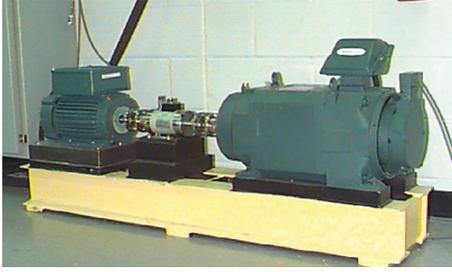


FIGURE 3: Experimental setup of bearing test rig in CWRU [34].

4.2. Discretization Effect Assessment. In the experiment, there are 12 columns of feature arrays and an array of corresponding bearing defect labels making up a 150×13 data matrix. The feature set is composed of eight features from time domain and four features from frequency domain, as listed in Table 1, and the types of bearing running states are total of 10 types. The matrix is processed using the equal probability-based discretization representation technique to divide each column feature array into 10 intervals. To validate the performance of the proposed discretization method intuitively, the interval class-information entropy is introduced. The results of the discretization methods based on equal probability, equal density, and equal width are listed in Table 2.

Besides, the RMS-based criteria have been used to compare with the newly interval class-information entropy criterion and evaluate the performance of the newly proposed criterion. The results of the discretization methods based on equal probability, equal density, and equal width are listed in Table 3.

Figure 4 is a line chart more clearly showing the performance comparison of different discretization methods. We can see that the average value of the interval class-information entropy and the RMS-based indicator based on equal probability are smaller than the other two methods, which indicates the proposed method is conspicuous with a better discretization performance. We draw distribution maps related to the equal probability discretization method and the two other discretization methods based on equal

density and equal width, as illustrated in Figure 5. Comparing with the other two methods, the waveform of the equal probability discretization method is most like the distribution of the raw data.

In addition, Support Vector Machine (SVM) is applied to discretization classification. There are 15 sets of data in each type of failure, including seven sets of training data and eight sets of predictions. Table 4 shows the results of classification accuracy of the three different discretization methods. The equal probability method also achieves the highest classification accuracy.

As a most basic discretization technique, evenly divided methods such as equal density and width approaches only consider the data boundary and neglect the overall distribution of the data, which facilitate excessive concentration or decentralized data, while the equal probability-based discretization method follows the guideline of equalized probability distribution of the data to naturally divide an array of features into reasonable sections, avoiding unbalancing each data extent and blindness caused by manually setting in traditional methods.

4.3. Representative Feature Mining of Bearing. In the experiment, the 150×13 matrix introduced previously is still used for further analysis. The number of the intervals is still 10, so the grade is divided into 10 levels. For easier observation and analysis, we transform each section into a glyph with sign like “ $a_1 a_2 \dots b_1 \dots e_{10}$.” The alphabet presents a specific feature, and the numeric as sign indicates the level of the interval. There are 12 column feature vectors, so “a-1” are used to represent features “ $f_{BPFO}, f_{BSF}, f_{BPFI}, f_{x_{MV}}, x_{SD}, x_{RMS}, x_{RA}, x_{CF}, x_s, x_K, x_{KC}$,” respectively. Thus, the numerical matrix is transformed into a symbolic matrix. Then, the symbolic matrix can be mined for generating a series of underlying association rules for bearing fault diagnosis by the Apriori algorithm [42]. The minimum support and confidence are set at 0.07 and 0.9, respectively. Table 5 presents the mining results.

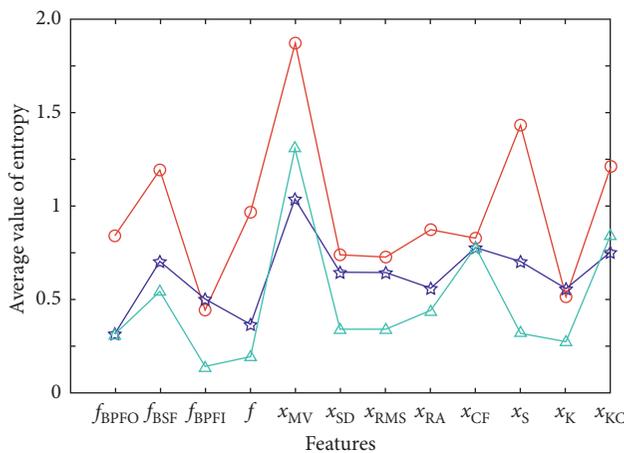
According to the mining results, a map of three bearing failures in different fault size is drawn to illustrate the range

TABLE 2: Interval class-information entropy based on different discretization methods.

	f_{BPFO}	f_{BSF}	f_{BPFI}	f	x_{MV}	x_{SD}	x_{RMS}	x_{RA}	x_{CF}	x_S	x_K	x_{KC}
Equal probability	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.80	0.00	1.43	1.45	1.45	1.50	1.99	0.00	0.00	0.00
	2.26	2.16	2.00	2.58	0.92	1.49	1.49	2.21	1.50	2.62	2.39	2.09
	2.23	1.47	0.00	1.00	2.13	1.13	1.13	1.46	0.95	1.53	1.00	1.05
	0.00	1.75	0.97	0.00	1.89	0.00	0.00	0.00	1.73	0.00	0.00	1.28
Equal density	0.00	0.00	0.00	0.35	1.56	0.00	0.00	0.00	0.00	1.77	0.00	0.91
	0.91	1.24	1.51	1.78	1.23	0.92	0.92	1.00	0.00	1.80	0.35	1.37
	1.37	1.46	1.43	1.81	2.15	1.23	1.23	1.29	1.23	1.05	1.38	0.92
	1.57	2.07	1.81	2.18	1.97	1.05	1.05	1.46	1.52	1.04	0.35	0.91
	1.51	1.43	0.84	2.09	2.15	1.81	1.89	1.69	1.23	1.55	1.16	0.57
Equal width	2.86	2.52	2.32	2.58	0.00	1.94	1.94	1.65	2.21	2.71	2.81	2.92
	0.67	1.85	0.87	0.50	0.00	1.79	1.79	1.94	1.33	0.34	0.00	1.83
	0.00	1.29	1.00	0.88	1.52	1.16	1.16	0.92	1.73	0.00	1.00	1.00
	0.00	0.98	0.85	0.00	2.17	0.00	0.00	0.00	0.00	0.00	0.72	1.28
	0.00	0.00	0.00	0.00	2.28	0.00	0.00	0.00	0.99	0.00	0.92	0.00

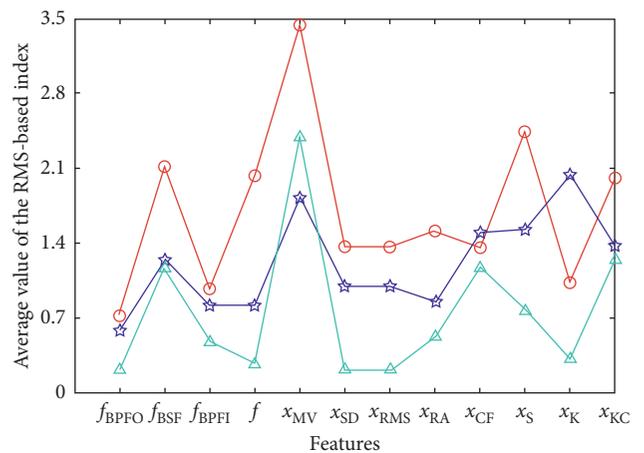
TABLE 3: The indicator of the RMS-based criteria based on different discretization methods.

	f_{BPFO}	f_{BSF}	f_{BPFI}	f	x_{MV}	x_{SD}	x_{RMS}	x_{RA}	x_{CF}	x_S	x_K	x_{KC}
Equal probability	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	1.00	2.06	0.00	3.09	2.78	2.78	2.83	4.23	0.00	0.00	0.00
	4.78	4.96	3.99	5.99	1.90	2.84	2.84	4.69	2.85	6.39	5.55	4.53
	4.82	2.81	1.00	2.00	4.49	2.32	2.32	2.78	1.94	3.46	2.00	2.25
	0.00	3.41	1.96	1.00	3.99	0.00	0.00	0.00	3.45	1.00	0.00	2.60
Equal density	1.00	1.00	1.00	1.51	3.22	1.00	1.00	1.00	1.00	3.68	1.00	2.13
	2.13	2.44	2.85	3.56	2.51	1.90	1.90	2.00	1.00	3.51	1.51	2.61
	2.61	2.75	2.75	3.61	4.52	2.51	2.51	2.61	2.51	2.27	2.85	1.90
	1.71	4.29	3.61	4.61	3.91	2.27	2.27	2.75	2.89	2.47	1.51	2.13
	1.65	2.75	1.81	4.37	4.52	3.61	3.61	3.38	2.51	3.08	2.40	1.62
Equal width	7.53	5.79	2.32	2.58	0.00	1.94	1.94	1.65	2.21	2.71	2.81	2.92
	1.68	3.65	0.87	0.50	0.00	1.79	1.79	1.94	1.33	0.34	0.00	1.83
	0.00	2.53	1.00	0.88	1.52	1.16	1.16	0.92	1.73	0.00	1.00	1.00
	0.00	0.98	0.85	0.00	2.17	0.00	0.00	0.00	0.00	0.00	0.72	1.28
	0.00	0.00	0.00	0.00	2.28	0.00	0.00	0.00	0.99	0.00	0.92	0.00



—○— Equal density
—☆— Equal width
—△— Equal probability

(a)



—○— Equal density
—☆— Equal width
—△— Equal probability

(b)

FIGURE 4: The performance comparison based on different discretization methods. (a) The average value of the interval class-information entropy. (b) The average value of the RMS-based indicator.

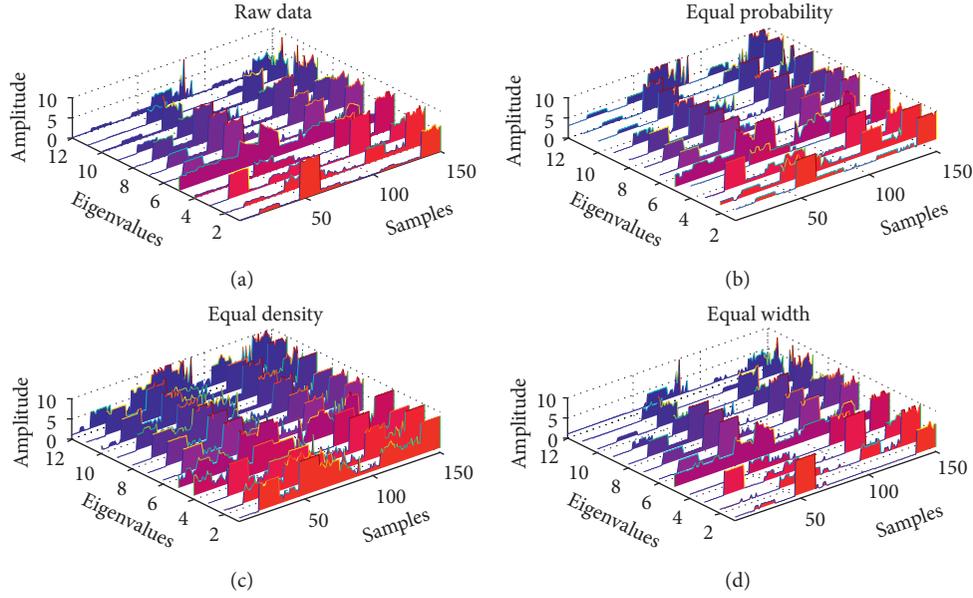


FIGURE 5: Performance comparison of features from time and frequency domains.

TABLE 4: Classification accuracy of SVM based on different discretization methods.

Discretization method	Equal density	Equal width	Equal probability
Classification accuracy	90%	93%	95%

TABLE 5: Results of mining association rules.

Types	f_{BPFO}	f_{BSF}	f_{BPFI}	f	x_{MV}	x_{SD}	x_{RMS}	x_{RA}	x_{CF}	x_S	x_K	x_{KC}
Normal	a_2		c_2	d_3	e_4	f_3	g_3	h_2	i_2	j_3	k_3	l_3
IR 0.007	a_4	b_4	c_9	d_3		f_6	g_6	h_7	i_5	j_4	k_4	l_4
IR 0.014	a_4			d_5		f_4	g_4	h_3			k_4	l_{10}
IR 0.021	a_4	b_8	c_{10}	d_{10}		f_9	g_9	h_{10}			k_6	l_5
B 0.007	a_3	b_3	c_3	d_3		f_3	g_3	h_3		j_3	k_3	l_3
B 0.014	a_3	b_3		d_3	e_2			h_2		j_3	k_3	
B 0.021		b_4	c_3			f_3	g_3	h_3		j_3	k_3	l_3
OR 0.007	a_{10}					f_{10}	g_{10}	h_{10}	i_9	j_{10}		l_5
OR 0.014	a_2	b_3	c_3	d_3		f_3	g_3		i_2	j_3	k_3	l_3
OR 0.021	a_{10}	b_{10}		d_{10}	e_2	f_{10}	g_{10}	h_8	i_{10}		k_{10}	

of characteristic value intuitively and it can clearly reflect the sensitivity and relevance between the features and bearing defects with the fault degree changing as shown in Figure 6. The eigenvalues associated with the normal running state located in the low level interval. That is, the feature amplitude of the fault-free operation is relatively small. For roller fault, the eigenvalues are close to that of the normal state except that f_{BSF} and x_{RA} which have increased slightly. In addition, since the eigenvalues of the fault in 0.014 in. diameter have an unusual performance, the priority is given to the features of the faults in diameters of 0.007 and 0.021 in. as for the inner race defect, and the amplitudes of f_{BSF} , f_{BPFI} , x_{SD} , x_{RMS} , and x_{RA} are much higher than the normal, of which f_{BPFI} changes dramatically, while the other features f_{BSF} , x_{SD} , x_{RMS} , and x_{RA} change steadily. When the defect

severities increase, the amplitude increases obviously. So do x_K and x_{KC} , although the changes are not so obvious. Similarly, it is easy to find that the features f_{BPFO} , x_{SD} , x_{RMS} , x_{RA} , and x_{CF} are more sensitive to the outer race defect, while f_{BPFO} fluctuates remarkably and is extracted as a representative feature. It is also found that with the increase of the outer race defect severities, the related features fluctuate distinctly instead of increasing steadily. The eigenvalues f_{BSF} and x_{RA} are selected as ball rolling element defect-related features and f_{BSF} is chosen as a representative feature fluctuating regularly. Representative features are listed in Table 6 according to different bearing defects. Then the range of eigenvalues in 0.007 and 0.021 in. fault is mined, as presented in Table 7, which can be used to find the type and size of the fault based on the respective values of the features.

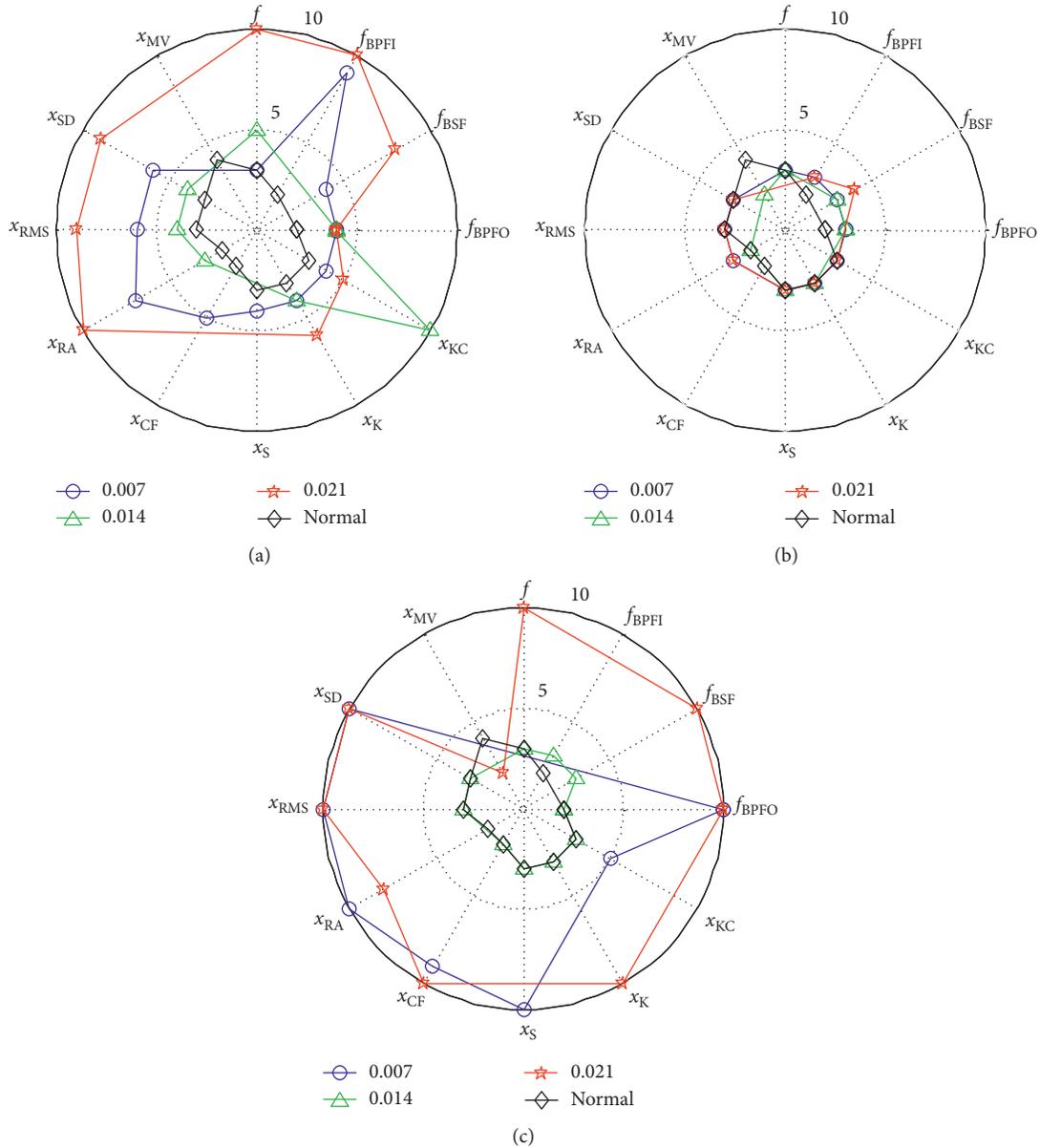


FIGURE 6: Three bearing failures in different fault sizes. (a) Inner race, (b) ball, and (c) outer race.

TABLE 6: Representative features of different defects.

Defect types	Related features		Representative features
	Time domain	Frequency domain	
Inner race	$x_{SD}, x_{RMS}, x_{RA}, x_K, x_{KC}$	f_{BSF}, f_{BPFI}	f_{BPFI}
Outer race	$x_{SD}, x_{RMS}, x_{RA}, x_{CF}$	f_{BPFO}	f_{BPFO}
Ball	x_{RA}	f_{BSF}	f_{BSF}

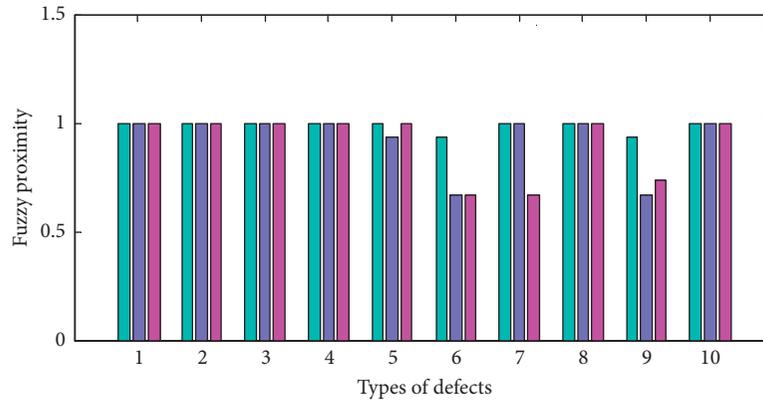
Fuzzy proximity is applied to validate the classification effectiveness of the rules mined by the proposed method, and the analysis results are shown in Figure 7. The fuzzy proximity classification accuracy of each type of bearing defect with the proposed method is generally higher than that of the other methods, and the average accuracy of the 10 types of bearing defects is calculated as shown in

Figure 7(b). The proposed method gets the highest accuracy of 98.67%, which demonstrates that the equal probability-based association rule mining is an effective and excellent method.

According to Table 7, nine representative features related to various defects can be selected from 12 features, and the nine selected features are used as the indicators of data

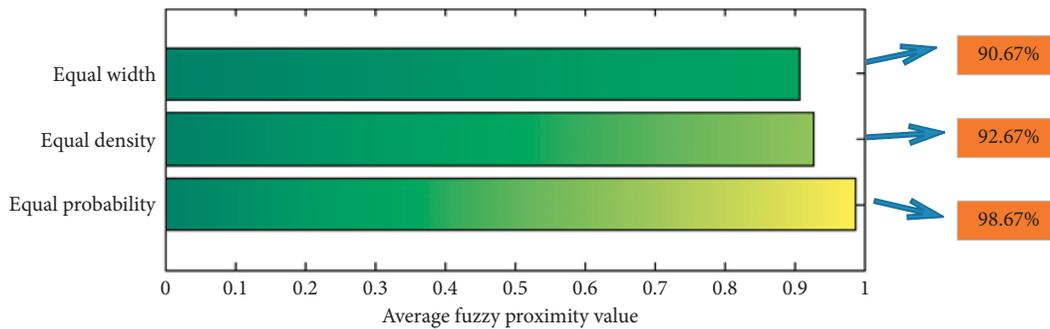
TABLE 7: The range of eigenvalues in different fault sizes.

Types	f_{BPFO}	f_{BSF}	f_{BPFI}	f	x_{MV}	x_{SD}	x_{RMS}	x_{RA}	x_{CF}	x_S	x_K	x_{KC}
Normal	-0.1124	—	-0.0289	-0.0011	0.0109	0.1053	0.1067	0.0310	-0.1860	-0.0042	-0.2403	2.4037
	-0.0351	—	0.0073	0.0330	0.0138	0.1736	0.1748	0.0682	0.5280	0.0033	0.0261	4.9411
IR 0.007	0.0211	0.0233	0.1455	-0.0011	—	0.2846	0.2856	0.1606	1.4855	0.0033	0.0261	4.9411
	—	—	—	—	—	—	—	—	—	—	—	—
IR 0.021	0.0686	0.0406	0.1817	0.0330	—	0.3379	0.3388	0.1834	1.8912	0.0095	0.2509	7.0821
	0.0211	0.0902	0.1817	0.2249	—	0.4639	0.4644	0.2478	—	—	0.4591	7.0821
OR 0.007	—	—	—	—	—	—	—	—	—	—	—	—
	0.0686	0.1108	0.2762	0.3528	—	0.5578	0.5581	0.2999	—	—	0.6672	9.0645
OR 0.021	0.3374	—	—	—	—	0.5578	0.5581	0.2478	3.2544	0.0450	—	7.0821
	—	—	—	—	—	—	—	—	—	—	—	—
OR 0.007	0.5577	—	—	—	—	0.6974	0.6978	0.2999	3.9685	0.0884	—	9.0645
	0.3374	0.1391	—	0.2249	0.0028	0.5578	0.5581	0.1834	3.9685	—	1.5247	—
OR 0.021	—	—	—	—	—	—	—	—	—	—	—	—
	0.5577	0.2539	—	0.3528	0.0075	0.6974	0.6978	0.2106	6.6529	—	4.4361	—



Equal probability
 Equal density
 Equal width

(a)



(b)

FIGURE 7: The classification accuracies of defect types using fuzzy proximity method. (a) The classification accuracy of each type of bearing defect and (b) the average accuracy of 10 types of bearing defects.

classification. SVM and BP neural networks are used as the classifier, respectively. Then, based on the three types of fault datasets including outer race fault, inner race fault, and rolling element faults, the traditional dimensionality reduction methods including PCA, KPCA, and LLE are chosen

to be contrast methods for reducing the data dimension from 12 to 9. The classification results are shown in Table 8. The representative features selected by the proposed method achieve superior performance compared to the traditional feature selection methods based on data transformation.

TABLE 8: Performance comparison of fault classification with different feature selection schemes.

Classifier Typical defects	SVM			BP		
	Inner race faults	Outer race faults	Rolling element faults	Inner race faults	Outer race faults	Rolling element faults
PCA	90	92.67	88.67	64.6	67.3	66.9
KPCA	72.67	71.44	66.2	56.4	58.5	52.67
LLE	69	64.8	67.1	29.92	33.4	32.52
Representative features	94.33	96.71	96	68.87	72.67	66.67

5. Conclusions

To improve the reliability of rotary machinery, effective and efficient diagnosis methods are highly needed. There is a new equal probability-based association rule mining method presented in this paper which provides an approach directly unearthing the underlying relation between labeled defects and unusual features for bearing fault analysis. In view of the shortcomings of the traditional evenly dividing methods used in association rule mining, this paper presents a new association rule mining approach based on the equal probability discretization method to avoid data excessive concentration or dispersion. First, a series of extracted features of signal data are discretized following the guideline of the equalized probability distribution of the data. Then, the data matrix composed of arrays of discretized features and defect labels is exploited to generate the association rules representing the relation between features and fault types. The rules are used for bearing fault diagnosis and help take a bearing defect signature analysis in a systematic manner. The proposed method has been compared with two other evenly dividing discretization methods in the experimental study. Moreover, as a new feature selection method, it does not need to transform the data matrix leading to bereave of the actual physical meaning of the selected features compared to the traditional PCA, KPCA, and LLE dimension reduction methods. From the analysis and compare results, conclusions can be drawn as follows:

- (1) Discretization is the most important process of the quantitative association rule mining. Two types of methods, including interval class-information entropy criteria and SVM, are used to assess discretization performance. It turns out that the equal probability method possesses prominent superiority.
- (2) From the study, some features, which can be called representative features, show their sensitivity to the bearing defects such as the ball pass frequency, and inner race to the bearing inner fault. And the feature map drew by the proposed method intuitively illustrates the sensitivity and relevance between the features and bearing defects as the fault degree changes. However, there are also some special circumstances, such as the mining result of bearing rolling element defect. There is no obvious characteristic except the inconspicuous relationship between ball (roller) spin frequency and root amplitude, which is partially due to the complex rolling mechanism. So, the proposed method also provides a new idea for feature selection which does not need to transform the data matrix leading to bereave of the actual physical meaning of the selected features.
- (3) The effectiveness of the proposed method is confirmed by the fuzzy proximity method compared to common evenly dividing discretization methods. The presented method gets the highest classification accuracy. In addition, the proposed method also achieves superior performance compared to the traditional feature selection methods based on data transformation.

Nomenclature

- x_i : Value of the i^{th} point in the sequence
 n : Length of time series
 f_r : Shaft speed
 N : Number of rolling elements
 D : Bearing pitch diameter
 d : Rolling element diameter
 ϕ : Angle of the load from the radial plane
 k : Number of intervals
 r_i : Total number of samples in the i^{th} interval
 c_{ij} : Number of samples in the i^{th} interval with the j^{th} eigenvalue.

Data Availability

The data used to support the findings of this study are acquired from the bearing data center of CWRU and the web page <http://csegroups.case.edu/bearingdatacenter/home> (accessed March 2019).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research acknowledges the financial support provided by the National Key Research and Development Program of China (no. 2016YFC0802103), National Science Foundation of China (nos. U1862104 and 51674277), and Science Foundation of China University of Petroleum, Beijing (no. ZX20180008).

References

- [1] F.-J. Ebert, "An overview of performance characteristics, experiences and trends of aerospace engine bearings technologies," *Chinese Journal of Aeronautics*, vol. 20, no. 4, pp. 378–384, 2007.
- [2] S. Deng, B. Jing, S. Sheng, Y. Huang, and H. Zhou, "Impulse feature extraction method for machinery fault detection using fusion sparse coding and online dictionary learning," *Chinese Journal of Aeronautics*, vol. 28, no. 2, pp. 488–498, 2015.
- [3] Z. Zhou, J. Zhao, and F. Cao, "A novel approach for fault diagnosis of induction motor with invariant character vectors," *Information Sciences*, vol. 281, pp. 496–506, 2014.
- [4] Y.-L. He, R. Wang, S. Kwong, and X.-Z. Wang, "Bayesian classifiers based on probability density estimation and their applications to simultaneous fault diagnosis," *Information Sciences*, vol. 259, pp. 252–268, 2014.
- [5] J. Liu, "Shannon wavelet spectrum analysis on truncated vibration signals for machine incipient fault detection," *Measurement Science and Technology*, vol. 23, no. 5, pp. 1–11, 2012.
- [6] J. Liu, W. Wang, and F. Ma, "Bearing system health condition monitoring using a wavelet cross-spectrum analysis technique," *Journal of Vibration and Control*, vol. 18, no. 7, pp. 953–963, 2012.
- [7] J. Chen, J. Pan, Z. Li, Y. Zi, and X. Chen, "Generator bearing fault diagnosis for wind turbine via empirical wavelet transform using measured vibration signals," *Renewable Energy*, vol. 89, pp. 80–92, 2016.
- [8] W. Guo, P. W. Tse, and A. Djordjević, "Faulty bearing signal recovery from large noise using a hybrid method based on spectral kurtosis and ensemble empirical mode decomposition," *Measurement*, vol. 45, no. 5, pp. 1308–1322, 2012.
- [9] Y. Guo, T.-W. Liu, J. Na, and R.-F. Fung, "Envelope order tracking for fault detection in rolling element bearings," *Journal of Sound and Vibration*, vol. 331, no. 25, pp. 5644–5654, 2012.
- [10] Y. Lei, J. Lin, Z. He, and Y. Zi, "Application of an improved kurtogram method for fault diagnosis of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 25, no. 5, pp. 1738–1749, 2011.
- [11] L. Wang, L. Cui, D. Zheng, and L. Gu, "Nonlinear dynamics behaviors of a rotor roller bearing system with radial clearances and waviness considered," *Chinese Journal of Aeronautics*, vol. 21, pp. 86–96, 2008.
- [12] S. V. Nalinaksh and D. Satishkumar, "Artificial neural network design for fault identification in a rotor-bearing system," *Mechanism and Machine Theory*, vol. 36, no. 2, pp. 157–175, 2001.
- [13] H. Wang and P. Chen, "Intelligent diagnosis method for rolling element bearing faults using possibility theory and neural network," *Computers & Industrial Engineering*, vol. 60, no. 4, pp. 511–518, 2011.
- [14] X. Zhang, W. Chen, B. Wang, and X. Chen, "Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization," *Neurocomputing*, vol. 167, pp. 260–279, 2015.
- [15] K. C. Gryllias and I. A. Antoniadis, "A support vector machine approach based on physical model training for rolling element bearing fault detection in industrial environments," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 2, pp. 326–344, 2012.
- [16] P. Konar and P. Chattopadhyay, "Bearing fault detection of induction motor using wavelet and support vector machines (SVMs)," *Applied Soft Computing*, vol. 11, no. 6, pp. 4203–4211, 2011.
- [17] Y. Pan, J. Chen, and X. Li, "Bearing performance degradation assessment based on lifting wavelet packet decomposition and fuzzy *c*-means," *Mechanical Systems and Signal Processing*, vol. 24, no. 2, pp. 559–566, 2010.
- [18] C. T. Yiakopoulos, K. C. Gryllias, and I. A. Antoniadis, "Rolling element bearing fault detection in industrial environments based on a *K*-means clustering approach," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2888–2911, 2011.
- [19] T. B. Bai, Y. L. Zhang, X. D. Wang, L. X. Duan, and J. J. Wang, "Bearing defect signature analysis based on a SAX-based association rule mining," in *proceedings of the IEEE International Conference on Prognostics and Health Management*, pp. 1–7, Ottawa, ON, Canada, June 2016.
- [20] W. Lin, S. A. Alvarez, and C. Ruiz, "Efficient adaptive-support association rule mining for recommender systems," *Data Mining and Knowledge Discovery*, vol. 6, no. 1, pp. 83–105, 2002.
- [21] J. Jiao and Y. Zhang, "Product portfolio identification based on association rule mining," *Computer-Aided Design*, vol. 37, no. 2, pp. 149–172, 2005.
- [22] X. Zheng and S. Wang, "Study on the method of road transport management information data mining based on pruning eclat algorithm and mapreduce," *Procedia—Social and Behavioral Sciences*, vol. 138, pp. 757–766, 2014.
- [23] M. Versichele, L. de Groote, M. C. Bouuaert, T. Neutens, I. Moerman, and N. Van de Weghe, "Pattern mining in tourist attraction visits through association rule learning on bluetooth tracking data: a case study of Ghent, Belgium," *Tourism Management*, vol. 44, pp. 67–81, 2014.
- [24] F. Kargarfard, A. Sami, and E. Ebrahimie, "Knowledge discovery and sequence-based prediction of pandemic influenza using an integrated classification and association rule mining (CBA) algorithm," *Journal of Biomedical Informatics*, vol. 57, pp. 181–188, 2015.
- [25] I. Vladimir, T. Ivan, T. Jasmina, M. Knežević, S. Elheshk, and I. Luković, "Using association rule mining to identify risk factors for early childhood caries," *Computer Methods and Programs in Biomedicine*, vol. 122, no. 2, pp. 175–181, 2015.
- [26] Z. Yang, W. H. Tang, A. Shintemirov, and Q. H. Wu, "Association rule mining-based dissolved gas analysis for fault diagnosis of power transformers," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 39, no. 6, pp. 597–610, 2009.
- [27] T.-Y. Li and X.-M. Li, "Preprocessing expert system for mining association rules in telecommunication networks," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1709–1715, 2011.
- [28] Y.-C. Hu, R.-S. Chen, and G.-H. Tzeng, "Discovering fuzzy association rules using fuzzy partition methods," *Knowledge-Based Systems*, vol. 16, no. 3, pp. 137–147, 2003.
- [29] Z. Hu, Z. Chen, W. Gui, and B. Jiang, "Adaptive PCA based fault diagnosis scheme in imperial smelting process," *ISA Transactions*, vol. 53, no. 5, pp. 1446–1455, 2014.
- [30] Y. Zhang and C. Ma, "Fault diagnosis of nonlinear processes using multiscale KPCA and multiscale KPLS," *Chemical Engineering Science*, vol. 66, no. 1, pp. 64–72, 2011.
- [31] Y. Liu, Z. Yu, M. Zeng, and Y. Zhang, "An improved LLE algorithm based on iterative shrinkage for machinery fault diagnosis," *Measurement*, vol. 77, pp. 246–256, 2016.

- [32] Z. Yin and J. Hou, "Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes," *Neurocomputing*, vol. 174, pp. 643–650, 2016.
- [33] T. Waqar and M. Demetgul, "Thermal analysis MLP neural network based fault diagnosis on worm gears," *Measurement*, vol. 86, pp. 56–66, 2016.
- [34] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 207–216, Washington, DC, USA, May 1993.
- [35] K.-I. Ahn, "Effective product assignment based on association rule mining in retail," *Expert Systems with Applications*, vol. 39, no. 16, pp. 12551–12556, 2012.
- [36] J. Nahar, T. Imam, K. S. Tickle, and Y.-P. P. Chen, "Association rule mining to detect factors which contribute to heart disease in males and females," *Expert Systems with Applications*, vol. 40, no. 4, pp. 1086–1093, 2013.
- [37] K. F. Bernard, F. Rigal, and F. Mauget, "Mining association rules for the quality improvement of the production process," *Expert Systems with Applications*, vol. 40, no. 4, pp. 1034–1045, 2013.
- [38] T. Li and X. Li, "Novel alarm correlation analysis system based on association rules mining in telecommunication networks," *Information Sciences*, vol. 180, no. 16, pp. 2960–2978, 2010.
- [39] R. Srikant and R. Agrawal, "Mining quantitative association rules in large relational tables," in *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data (SIGMOD '96)*, pp. 1–12, Montreal, Canada, June 1996.
- [40] E. Keogh and J. Lin, "Symbolic Aggregate approXimation (SAX)," August 2011, <http://cs.gmu.edu/~jessica/sax.htm>.
- [41] Case Western Reserve University bearing data center website, <http://csegroups.case.edu/bearingdatacenter/pages/apparatus-procedures>.
- [42] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proceedings of the 20th International Conference on Very Large Data Bases*, pp. 478–499, Santiago, Chile, September 1994.

