Springer

KSME

Journal of Mechanical Science and Technology 32 (11) (2018) 5201~5211
www.springerlink.com/content/1738-494x(Print)/1976-3824(Online)
DOI 10.1007/s12206-018-1018-7

Rolling bearing fault diagnosis based on mean multigranulation decision-theoretic rough set and non-naive Bayesian classifier†

Jun Yu[1], Bo Ding[2] and Yongjun He[2,*]

[1]*School of Automation, Harbin University of Science and Technology, Harbin 150080, China*
[2]*School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China*

(Manuscript Received March 25, 2018; Revised July 2, 2018; Accepted July 15, 2018)

Abstract

To analyze data from multi-level view, reduce computational burden, and improve fault diagnosis accuracy, a novel fault diagnosis method of rolling bearings based on mean multigranulation decision-theoretic rough set (MMG-DTRS) and non-naive Bayesian classifier (NNBC) is proposed in this paper. First, fault diagnosis features of rolling bearings in training samples are extracted to construct MMG-DTRS. Then, the significance degree of condition attribute in MMG-DTRS is defined to quantitatively measure the influence of condition attributes with respect to the decision ability of an information system. An attribute reduction algorithm based on MMG-DTRS is applied to acquire a lower dimensional condition attribute set, which reduces computational complexity and avoids the interference of irrelevant or redundant condition attributes. Finally, NNBC is constructed to classify rolling bearing conditions in test samples. The classification procedures by using NNBC are given. The performance of the proposed method is validated and the advantages are investigated by using a fault diagnosis experiment of rolling bearings. Experimental investigations demonstrate the proposed method is effective and reliable in identifying fault categories and fault severities of rolling bearings.

Keywords: Rolling bearing; Fault diagnosis; Mean multigranulation decision-theoretic rough set; Non-naive Bayesian classifier; Attribute reduction

---

1. Introduction

Rolling bearings are one of the key components in rotating machinery, such as wind turbines, motors, and planetary gearboxes, etc. [1-3]. Due to harsh operating conditions caused by heavy load, high speed, or contamination, rolling bearings are prone to various faults such as crack, pitting and spalling, etc. [4, 5]. Once they suffer from the faults, they may lead to entire system shutdown, huge production losses or even catastrophic results. Therefore, fault diagnosis is a crucial task to guarantee healthy operation state of rolling bearings.

Recently, various data-driven fault diagnosis methods of rolling bearings have been put forward. Many intelligent techniques, such as deep neural network (DNN), support vector machine (SVM), and fuzzy logic (FL), have been investigated and developed as novel tools for rolling bearing fault diagnosis [6]. Deep learning as a novel machine learning tool possesses the capability to overcome the inherent disadvantages of traditional intelligent techniques [7, 8]. The most obvious advantage of deep learning models lies in that they can automatically discovery valuable information from raw data. The most popular deep learning model is DNN, which has been extensively applied in rolling bearing fault diagnosis owing to its strong representation ability and simple structure [9-13]. Gan et al. [9] put forward a hierarchical diagnosis network (HDN) using deep learning. A two stage diagnosis strategy is used for hierarchical classification through a two-layer HDN. The first layer is utilized to classify fault types, and the second one is employed to realize fault severity reorganization. Guo et al. [10] proposed an adaptive hierarchical deep convolution neural network for fault type recognition and fault size evaluation of rolling bearings. Shao et al. [11] used ensemble deep auto-encoders (EDAEs) to propose an intelligent fault diagnosis method of rolling bearings. Different activation functions are regarded as the hidden functions of a series of auto-encoders. EDAEs are constructed by these auto-encoders through unsupervised feature learning. Then, these learned deep features are considered as the input of classifiers for fault identification. Shao et al. [12] presented a fault diagnosis method of electric locomotive bearings using convolutional deep belief network (CDBN). Collected vibration data is compressed through an auto-encoder. A CDBN constructed through Gaussian visible units is used to learn deep features. The constructed CDBN is improved by an exponential moving average. Zhang et al. [13] constructed a deep convolu-

*Corresponding author. Tel.: +86 451 86390650
E-mail address: wanghai121222@163.com
† Recommended by Associate Editor Doo Ho Lee

tional neural network model for bearing fault diagnosis under variable working load and noisy environment. This model can achieve very high accuracy without any domain adaptation algorithm or any target domain information. However, determination of DNNs' structure and parameters requires a great number of training samples. High computational cost and low adaptation ability remains unsolved.

SVM, as a supervised machine learning technique, has been widely applied in fault diagnosis owing to its high prediction accuracy and outstanding generalization ability [14-19]. Besides, it does not need a large number of training samples to realize mode identification. Li et al. [15] combined hierarchical fuzzy entropy (HFE) with improved SVM to propose a bearing fault diagnosis method. HFE is utilized to extract fault features. Then, refined fault features are regarded as the input of the improved SVM to automatically identify fault patterns. Liu et al. [16] constructed a dictionary using roller bearing mechanism to put forward a diagnose method combining short-time matching and SVM. Practical bearing experimental results indicate the proposed method can classify a weak fault at the early stage from complex and non-stationary signals. Li et al. [17] presented a bearing fault diagnosis approach combining multiscale permutation entropy with improved SVM. Experimental results demonstrate this approach has high classification accuracy. Ziani et al. [18] proposed a bearing fault diagnosis strategy combining SVM with binary particle swarm optimization (BPSO). Regularized Fisher's criterion is considered as the fitness function of BPSO algorithm. Zheng et al. [19] put forward a fault detection and diagnosis method of rolling bearings combining composite multiscale fuzzy entropy (CMFE) with ensemble SVM. CMFE is used to measure the complexity of rolling bearing signals and extract fault features. The ensemble SVM is used for fault identification. Nevertheless, the parameter selection of SVM is a time-consuming process and many parameter selection algorithms are not satisfactory.

FL technique transforms qualitative, vague or ambiguous data into numerical terms. Ambiguous judgments can be realized according to uncertain or imprecise information. Consequently, FL technique has been successfully used in rolling bearing fault diagnosis [20-25]. Straczkiewicz et al. [20] presented a fault diagnosis approach of rolling element bearings using FL technique for integration of vibration-based diagnostic features. Ziani et al. [21] combined multi-scale permutation entropy (MPE) and adaptive neuro fuzzy classifier (ANFC) to propose a bearing fault diagnosis method. MPE is used to extract fault features and refine these features. The refined features are considered as the input of ANFC to classify fault modes. Xu et al. [22] put forward an intelligent fault diagnosis approach using selective ensemble of multiple fuzzy classifiers. Optimal features are regarded as the input of the selective ensemble of multiple fuzzy classifiers to classify fault patterns. Sun et al. [23] presented a fault diagnosis strategy of rolling bearings integrating fuzzy evidence discovery with Dempster–Shafer evidence theory. Li et al. [24] introduced a probabilistic fuzzy system to classify bearing faults. Each rule in the rule-based system can determine a fault and assign a probability to this fault. Experimental analyses demonstrate that this fuzzy system can match the effectiveness of other intelligent diagnosis methods. Although FL technique can achieve satisfactory diagnostic results by fuzzy decision rules, the procedures of acquiring fuzzy decision rules are very time-consuming. Moreover, construction of fuzzy system relies on human expertise and many parameter estimation algorithms are quite complicated.

In recent years, multigranulation rough sets (MGRSs) have attracted considerable attention for their multi-level view [26, 27]. They describe lower and upper approximations via multiple granulation relations to realize approximate approximation. Different from traditional rough sets, MGRSs are constructed according to a series of equivalence relations rather than a single one. Therefore, they have been successfully applied in attribute reduction and fault diagnosis fields. Qian et al. [28] introduced an importance measure method of condition attributes to realize attribute reduction based on pessimistic MGRS model. Tan et al. [29] developed belief reduction algorithms to find pessimistic lower (or upper) approximate reduct by evidence theory-based characteristics. Several measurements by using the belief and plausibility functions are introduced, and attribute reduction algorithms based on these measurements are given. Zhang et al. [30] proposed a fault diagnosis approach of steam turbine based on single-valued neutrosophic MGRSs over two universes. Single-valued neutrosophic MGRSs over two universes are defined. Then, general decision rules are constructed using rough sets over two universes for fault diagnosis. Zhang et al. [31] combined interval-valued hesitant fuzzy sets with MGRS to present a novel fault diagnosis method. Experimental results demonstrate that this method can increase fault diagnosis accuracy and greatly reduce uncertainty. During the attribute reduction using MGRSs, seeking common ground while removing differences is the most frequently used reduction strategy. It can be considered as a conservative reduction strategy, which means that it retains common rules while removing inconsistent rules. Hence, most of the attribute reduction algorithms by using MGRSs are based on pessimistic MGRS model. However, compared to optimistic MGRS model, constraint conditions of pessimistic MGRS model are excessively harsh. Sensitive and essential condition attributes are probably removed and fault diagnosis accuracy is reduced in practical application.

To overcome this problem, based on mean multigranulation decision-theoretic rough set (MMG-DTRS) and non-naive Bayesian classifier (NNBC), a novel fault diagnosis method of rolling bearings is put forward in this paper. First, fault diagnosis features of rolling bearings in training samples are extracted to construct MMG-DTRS. Then, an attribute reduction algorithm based on MMG-DTRS is applied to acquire a lower dimensional condition attribute set. Finally, NNBC is constructed to classify rolling bearing conditions in test samples. Experimental results demonstrate that this method can effectively and accurately identify fault categories and fault severities of rolling bearings.

The remainder of this paper is organized as follows. Sec. 2 briefly introduces the preliminaries of MGRSs, MMG-DTRS and NNBC. A fault diagnosis method of rolling bearings based on MMG-DTRS and NNBC is proposed in Sec. 3. The experimental results are analyzed and discussed in Sec. 4. Finally, conclusions are drawn in Sec. 5.

2. Preliminaries

2.1 Multigranulation rough sets

MGRSs were introduced by Qian et al. [26, 27]. Different from traditional rough sets, MGRSs are constructed according to a series of equivalence relations rather than a single one. Some basic concepts about MGRSs are summarized as follows.

An information system (IS) is defined as a quadruple (U,A,V,f), where U is a non-empty finite set of cases called the universe, $A=C\bigcup D$ is a non-empty finite set of attributes, where C is the condition attribute set and D is the decision attribute. V is regarded as the finite set of attributes. Each attribute $a\in A$ is associated with a set V_a, where V_a is the set of values of a, called the domain of attribute a. $f:U\times A\to V_a$ is a total function such as $f(x,a)\in V_a$ for each $a\in A$ and $x\in U$.

Definition 1 [26]. Let $(U,A=C\bigcup D,V,f)$ be an IS in which $B_1, B_2,..., B_m\subseteq C$, then $\forall X\subseteq U$, the optimistic lower approximation and upper approximation of X are denoted by $\underline{\sum_{i=1}^{m}B_i}^{O}(X)$ and $\overline{\sum_{i=1}^{m}B_i}^{O}(X)$, respectively,

$$\underline{\sum_{i=1}^{m}B_i}^{O}(X)=\{x\in U:[x]_{B_1}\subseteq X\vee[x]_{B_2}\subseteq X\vee...\vee[x]_{B_m}\subseteq X\} \quad (1)$$

$$\overline{\sum_{i=1}^{m}B_i}^{O}(X)=\sim\underline{\sum_{i=1}^{m}B_i}^{O}(\sim X) \quad (2)$$

where $[x]_{B_i}(1\le i\le m)$ are the equivalence classes of the case x with respect to the attribute subset B_i, and $\sim X$ is the complement set of X. $\left\langle \underline{\sum_{i=1}^{m}B_i}^{O}(X), \overline{\sum_{i=1}^{m}B_i}^{O}(X)\right\rangle$ is the optimistic MGRS model.

Proposition 1. Let $(U,A=C\bigcup D,V,f)$ be an IS in which $B_1, B_2,..., B_m\subseteq C$, then $\forall X\subseteq U$, we have

$$\overline{\sum_{i=1}^{m}B_i}^{O}(X)=\{x\in U:[x]_{B_1}\cap X\neq\varnothing\wedge[x]_{B_2}\cap X\neq\varnothing\wedge...\wedge[x]_{B_m}\cap X\neq\varnothing\}. \quad (3)$$

Definition 2 [26]. Let $(U,A=C\bigcup D,V,f)$ be an IS in which $B_1, B_2,..., B_m\subseteq C$, then $\forall X\subseteq U$, the pessimistic lower approximation and upper approximation of X are denoted by $\underline{\sum_{i=1}^{m}B_i}^{P}(X)$ and $\overline{\sum_{i=1}^{m}B_i}^{P}(X)$, respectively,

$$\underline{\sum_{i=1}^{m}B_i}^{P}(X)=\{x\in U:[x]_{B_1}\subseteq X\wedge[x]_{B_2}\subseteq X\wedge...\wedge[x]_{B_m}\subseteq X\} \quad (4)$$

$$\overline{\sum_{i=1}^{m}B_i}^{P}(X)=\sim\underline{\sum_{i=1}^{m}B_i}^{P}(\sim X) \quad (5)$$

$\left\langle \underline{\sum_{i=1}^{m}B_i}^{P}(X), \overline{\sum_{i=1}^{m}B_i}^{P}(X)\right\rangle$ is the pessimistic MGRS model.

Proposition 2. Let $(U,A=C\bigcup D,V,f)$ be an IS in which $B_1, B_2,..., B_m\subseteq C$, then $\forall X\subseteq U$, we have

$$\overline{\sum_{i=1}^{m}B_i}^{P}(X)=\{x\in U:[x]_{B_1}\cap X\neq\varnothing\vee[x]_{B_2}\cap X\neq\varnothing\vee...\vee[x]_{B_m}\cap X\neq\varnothing\}. \quad (6)$$

In the optimistic MGRS model, the word "optimistic" means that in multiple granular structures, it needs only at least a granular structure to satisfy with the inclusion condition between knowledge granule and target concept. In the pessimistic MGRS model, "pessimistic" means that in multiple granular structures, it needs all granular structures to satisfy with the inclusion condition between knowledge granule and target concept.

2.2 Mean multigranulation decision-theoretic rough set

Decision-theoretic rough sets (DTRSs) [32], as a form of probabilistic tools, possess very strong theoretical basis. DTRSs introduce Bayesian decision procedure to minimize decision costs. They provide a systematic strategy to set the threshold parameters according to loss functions. In the Bayesian decision procedure, a finite set of states can be written as $\Omega=\{X, \sim X\}$. A finite action set can be denoted by $A=\{a_P, a_N, a_B\}$, where a_P, a_N and a_B denote three actions, namely, deciding $x\in POS(X)$, deciding $x\in NEG(X)$, and deciding $x\in BND(X)$. λ_{11}, λ_{21} and λ_{31} represent the loss for taking actions of a_P, a_N and a_B, respectively, when a sample belongs to X. Similarly, λ_{12}, λ_{22} and λ_{32} represent the loss for taking the correspondence actions when the sample belongs to $\sim X$. For a sample x, the expected loss associated with taking individual actions can be expressed as:

$$R(a_1\mid[x])=\lambda_{11}P(X\mid[x])+\lambda_{12}P(\sim X\mid[x]),$$

$$R(a_2 \mid [x]) = \lambda_{21} P(X \mid [x]) + \lambda_{22} P(\sim X \mid [x]),$$
$$R(a_3 \mid [x]) = \lambda_{31} P(X \mid [x]) + \lambda_{32} P(\sim X \mid [x]).$$

The Bayesian decision procedure suggests the minimum-risk decision rules:

(P) If $P(X \mid [x]) \geq \gamma$ and $P(X \mid [x]) \geq \alpha$, then decision $x \in POS(X)$;

(N) If $P(X \mid [x]) \leq \beta$ and $P(X \mid [x]) \leq \gamma$, then decision $x \in NEG(X)$;

(B) If $\beta \leq P(X \mid [x]) \leq \alpha$, then decide $x \in BND(X)$;

where $\alpha = \dfrac{\lambda_{12} - \lambda_{32}}{(\lambda_{31} - \lambda_{32}) - (\lambda_{11} - \lambda_{12})}$, $\gamma = \dfrac{\lambda_{12} - \lambda_{22}}{(\lambda_{21} - \lambda_{22}) - (\lambda_{11} - \lambda_{12})}$, $\beta = \dfrac{\lambda_{32} - \lambda_{22}}{(\lambda_{21} - \lambda_{22}) - (\lambda_{31} - \lambda_{32})}$.

In many real applications, such as data analysis, attribute reduction, and mode identification of multi-source data with high dimensions, DTRS is not suitable for these cases. To overcome this issue, Qian et al. [33] combined MGRS and DTRS to propose MMG-DTRS. With respect to MGRSs, when the loss function is assigned, the conditional probability of a sample within a target concept in granular structures can be calculated as follows:

$$E(P(X \mid x)) = \{P(X \mid [x]_{B_1}) + P(X \mid [x]_{B_2}) + \ldots + P(X \mid [x]_{B_m})\} / m \tag{7}$$

where $[x]_{B_i}$ $(1 \leq i \leq m)$ is the equivalence class of x induced by B_i and $P(X \mid [x]_{B_i})$ is the conditional probability of the equivalent class $[x]_{B_i}$ with respect to X. The joint probability can be computed by the mean value of m conditional probabilities. According to this idea, MMG-DTRS can be defined as follows.

Definition 3 [33]. Let $(U, A = C \cup D, V, f)$ be an IS in which $B_1, B_2, \ldots, B_m \subseteq C$, then $\forall X \subseteq U$, the mean multigranulation decision-theoretic lower approximation, upper approximation, and boundary region of X are denoted by $\underline{\sum_{i=1}^{m} B_i}^{M,\alpha}(X)$, $\overline{\sum_{i=1}^{m} B_i}^{M,\beta}(X)$ and $BN^M_{\sum_{i=1}^{m} B_i}(X)$, respectively,

$$\underline{\sum_{i=1}^{m} B_i}^{M,\alpha}(X) = \{x \in U : \{P(X \mid [x]_{B_1}) + P(X \mid [x]_{B_2}) + \ldots + P(X \mid [x]_{B_m})\} / m \geq \alpha\} \tag{8}$$

$$\overline{\sum_{i=1}^{m} B_i}^{M,\beta}(X) = U - \{x \in U : \{P(X \mid [x]_{B_1}) + P(X \mid [x]_{B_2}) + \ldots + P(X \mid [x]_{B_m})\} / m \leq \beta\} \tag{9}$$

$$BN^M_{\sum_{i=1}^{m} B_i}(X) = \overline{\sum_{i=1}^{m} B_i}^{M,\beta}(X) - \underline{\sum_{i=1}^{m} B_i}^{M,\alpha}(X) \tag{10}$$

where α, β are two thresholds.

When the thresholds $\alpha > \beta$, the MMG-DTRS possesses the following decision rules:

(MP1) If $\{P(X \mid [x]_{B_1}) + P(X \mid [x]_{B_2}) + \ldots + P(X \mid [x]_{B_m})\} / m \geq \alpha$, then decision $x \in POS(X)$;

(MN1) If $\{P(X \mid [x]_{B_1}) + P(X \mid [x]_{B_2}) + \ldots + P(X \mid [x]_{B_m})\} / m \leq \beta$, then decision $x \in NEG(X)$;

(MB1) If $\beta < \{P(X \mid [x]_{B_1}) + P(X \mid [x]_{B_2}) + \ldots + P(X \mid [x]_{B_m})\} / m < \alpha$, then decide $x \in BND(X)$.

MMG-DTRS not only retains fault-tolerant ability of DTRS, but also realizes data analysis from multi-level view. Furthermore, it overcomes the shortcoming that constraint conditions of pessimistic MGRS model are excessively harsh. Thus, MMG-DTRS extends the wider applications such as data analysis and attribution reduction of multi-source information.

2.3 Non-naive Bayesian classifier

Recently, many methods have been developed for mode classification, such as ANN, SVM, and Dempster–Shafer theory (DST), etc. Nevertheless, two main disadvantages of the existing mode classification methods are summarized as follows. The huge computational burden prohibits the use of these methods in practice. Let N be the size of training samples, the training complexities of ANN, SVM and DST are $O(N^2)$, $O(N^3)$ and $O(N^2)$ respectively. Besides, they ignore the dependence among features in the acquired samples in practical applications. Bayesian classifiers are identification tools on the basis of statistical theory. Among Bayesian classifiers, naive Bayesian classifiers (NBCs) are a simple and efficient probabilistic tool, which have been successfully applied to identify fault modes [34]. The classification performance of NBCs is obviously superior to traditional mode identification tools due to their structure simplicity, high computational efficiency, and less storage requirement. A key step in the classification strategy of NBCs is to estimate probability density function (PDF) from training data set. However, the common methods of PDF estimation are on the basis of the assumption that all attributes are independent and this hypothesis is not suitable for many real-world applications [35]. To address this issue, NNBC using the optimal bandwidth selection was proposed to ignore the independence assumption among attributes and replace marginal PDF estimation by joint PDF estimation [36]. The class label of a sample is determined as the following:

$$w = \arg\max \left\{ \frac{n_k}{N} P(x \mid w_k) \right\} =$$

$$\arg\max\left\{\frac{1}{Nh_k^{(d)}}\sum_{i=1}^{n_k}\left[k\left(\frac{x_1-x_{i1}^{(k)}}{h_k},\ \frac{x_2-x_{i2}^{(k)}}{h_k},\ \ldots,\ \frac{x_d-x_{id}^{(k)}}{h_k},\right)\right]\right\} \tag{11}$$

where d is the attribute vector dimension and $N=\sum_{k=1}^{c} n_k$ where c is the class number. Besides, h_k is the optimal bandwidth for kth class. It can be computed by the following equation:

$$h_k=\left(\frac{4d}{n_k\left|\Sigma_k\right|^{-\frac{1}{2}}\left\{2tr\left(\Sigma_k^{-1}\Sigma_k^{-1}\right)+tr^2\left(\Sigma_k^{-1}\right)\right\}}\right)^{\frac{1}{d+4}} \tag{12}$$

where $\Sigma_k = diag\left\{\left[\sigma_1^{(k)}\right]^2,\ \left[\sigma_2^{(k)}\right]^2,\ \ldots,\ \left[\sigma_d^{(k)}\right]^2\right\}$.

The Gaussian kernel function $k(x)$ can be calculated by

$$k(x)=\frac{1}{\left(\sqrt{2\pi}\right)^d}\exp\left(-\frac{xx^T}{2}\right). \tag{13}$$

From Eq. (11), the class label of the sample x represents the probability that the sample x belongs to this class. The larger the class label is, the higher the probability that the sample x belongs to this class is. Accordingly, the class of the sample x can be determined through the maximum class label. Recent research demonstrates that NNBC can obtain the optimal classification performance among the existing NBCs. Moreover, it can achieve satisfactory diagnosis effectiveness without high time consumption and memory requirement.

3. Fault diagnosis method of rolling bearings based on MMG-DTRS and NNBC

3.1 Fault diagnosis framework

MMG-DTRS retains fault-tolerant ability of DTRS, and can realize data analysis from multi-level view. It also overcomes the shortcoming that constraint conditions of pessimistic MGRS model are excessively harsh. Moreover, NNBC based on the optimal bandwidth selection ignores the independence assumption among attributes and replaces marginal PDF estimation by joint PDF estimation. It can obtain satisfactory diagnosis accuracy without high time consumption and memory requirement. Therefore, we integrate MMG-DTRS with NNBC to propose a novel fault diagnosis method of rolling bearings, and the fault diagnosis framework is shown in Fig. 1. First, fault diagnosis features of rolling bearings in training samples are extracted to construct MMG-DTRS. Then, an attribute reduction algorithm based on MMG-DTRS is applied to acquire a lower dimensional condition attribute set. Finally, NNBC is constructed to classify rolling bearing conditions in test samples.

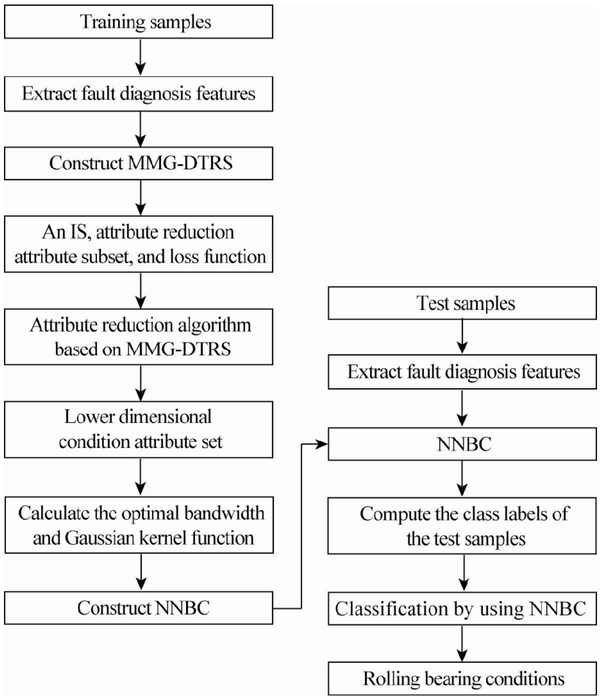


Fig. 1. Fault diagnosis framework of rolling bearings based on MMG-DTRS and NNBC.

3.2 Attribute reduction based on MMG-DTRS

Attribute reduction of condition attribute set is a vital topic in data-driven fault diagnosis methods. It removes irrelevant or redundant condition attributes and retains the decision ability of an IS. That is, the positive region determined by condition attributes after attribute reduction is consistent with the one before attribute reduction. Consequently, to quantitatively measure the influence of condition attributes with respect to the decision ability of an IS, the significance degree of condition attribute in MMG-DTRS can be defined as follows:

Definition 4. Let $(U, A=C\cup D, V, f)$ be an IS in which $B_1, B_2, \ldots, B_m \subseteq C$, then the classification quality of condition attribute set C with respect to decision attribute D in MMG-DTRS is defined as

$$S_C(D)=\frac{\sum\left\{X\in U/\{D\}:\underline{\overset{M,\alpha}{\sum_{i=1}^{m}B_i}}(X)\right\}}{|U|}. \tag{14}$$

Definition 5. Let $(U, A=C\cup D, V, f)$ be an IS in which $B_1, B_2, \ldots, B_m \subseteq C$ and condition attribute set $C=\{c_1, c_2, \ldots, c_n\}$, then $\forall c_i\in C(1\le i\le n)$, the significance degree of condition attribute c_i with respect to decision attribute D in MMG-DTRS is defined as

$$Sig(c_i, C, D)=S_C(D)-S_{C-\{c_i\}}(D). \tag{15}$$

Definition 6. Attribute c_i in condition attribute set C with respect to decision attribute D in MMG-DTRS is not the lower approximation significant if and only if $Sig(c_i, C, D) = 0$.

MMG-DTRS retains fault-tolerant ability of DTRS, and overcomes the shortcoming of harsh constraint conditions of pessimistic MGRS model. Furthermore, the significance degree of condition attribute in MMG-DTRS can be used to quantitatively measure the influence of condition attribute with respect to the decision ability of an IS. Therefore, an attribute reduction algorithm based on MMG-DTRS is proposed to acquire a lower dimensional condition attribute set. The flowchart of the proposed algorithm is shown in Fig. 2, and the general procedures are summarized as follows.

Input: An $IS = (U, A = C \cup D, V, f)$, an attribute subset $B_1, B_2, ..., B_m \subseteq C$, and loss function $\lambda_{p,q}(p, q = 1,2,3)$.

Output: A lower dimensional condition attribute set.

Step 1: Calculate the threshold α according to the loss function $\lambda_{p,q}(p, q = 1,2,3)$;

Step 2: Compute the classification quality $S_C(D)$ of condition attribute set C with respect to decision attribute D in MMG-DTRS;

Step 3: Calculate the significance degree $Sig(c_1, C, D)$ of condition attribute c_1 with respect to decision attribute D in MMG-DTRS;

Step 4: If the significance degree $Sig(c_i, C, D) = 0$, then the condition attribute c_1 is redundant, or else the condition attribute c_1 is indispensable;

Step 5: Repeat steps 3 and 4 to the other condition attributes, till the last one;

Step 6: Remove all the redundant condition attributes and generate a lower dimensional condition attribute set.

3.3 Classification by using NNBC

From the definitions of NNBC, it can be observed that NNBC based on the optimal bandwidth selection ignores the independence assumption among attributes and replaces marginal PDF estimation by joint PDF estimation. From Eq. (5), it can be noted that the larger the class label is, the higher the probability that the sample x belongs to this class is. The class of a sample x can be determined through the maximum class label. Therefore, NNBC is used to classify rolling bearing conditions in test samples in this paper. The classification procedures of a test sample x by using NNBC are summarized as follows.

Input: A test sample x and training samples after attribute reduction.

Output: The class of the test sample x.

Step 1: The optimal bandwidth h_k for kth class and the Gaussian kernel function $k(x)$ are representatively calculated according to training samples by Eqs. (12) and (13);

Step 2: Compute the class labels of the test sample x by Eq. (11);

Step 3: Determine the class of the test sample x according to the maximum class label.

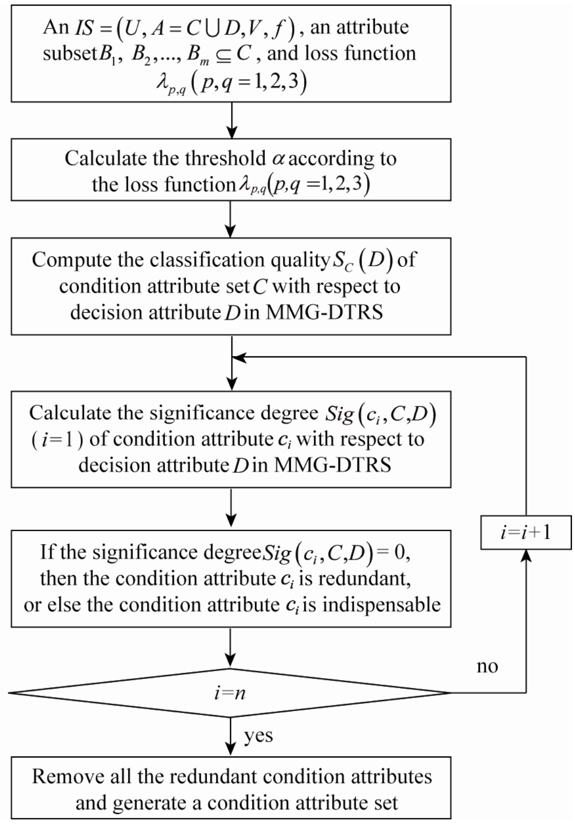


Fig. 2. Flowchart of the attribute reduction algorithm based on MMG-DTRS.

In the process of classification by NNBC, each class of the test sample x represents a rolling bearing condition. The class label of a sample x represents the probability that the sample x belongs to corresponding rolling bearing condition. The larger the class label is, the higher the probability that the sample x belongs to corresponding rolling bearing condition is. Consequently, the rolling bearing condition of the test sample x is determined through the maximum class label.

4. Experimental investigations

4.1 Experimental setup

A rolling bearing test-rig was used to evaluate the proposed method. The rolling bearing test-rig is illustrated in Fig. 3. A variable speed motor is employed to drive a rotating shaft supported by two rolling bearings. The rotating shaft is connected with the drive motor by a coupling. The strategy reduces the transmission and misalignment influences. Three weights are used to apply radial load. One SR150M acoustic emission (AE) sensor is mounted on the bearing housing to acquire AE signals. In the experiment, there are seven rolling bearing conditions: normal condition (NC), an inner race fault with a fault sizes of 0.4 mm (IRF-0.4 mm), an inner race fault with a fault sizes of 0.8 mm (IRF-0.8 mm), an outer race fault with a fault sizes of 0.4 mm (ORF-0.4 mm), an outer race fault with a fault sizes of 0.8 mm (ORF-0.8 mm), and a rolling

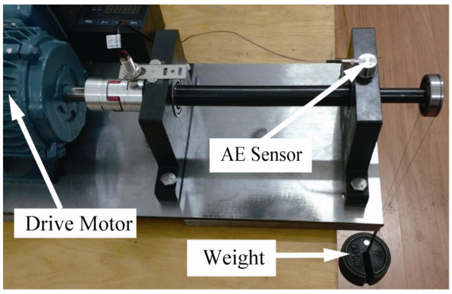


Fig. 3. Rolling bearing test-rig.

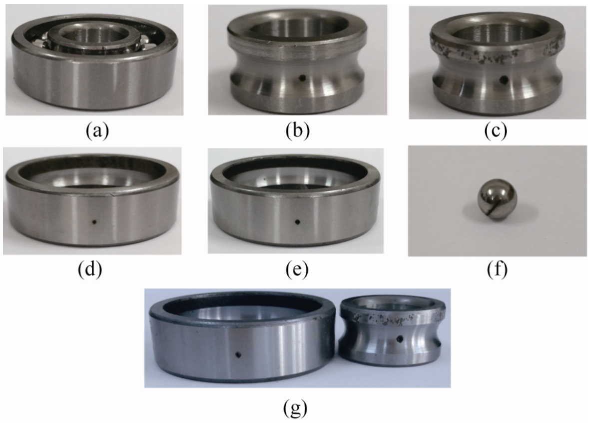


Fig. 4. Seven rolling bearing conditions: (a) NC; (b) IRF-0.4 mm; (c) IRF-0.8 mm; (d) ORF-0.4 mm; (e) ORF-0.8 mm; (f) REF; (g) CF.

element fault (REF), and combined fault (CF: IRF-0.8 mm and ORF-0.8 mm), respectively. Seven rolling bearing conditions are illustrated in Fig. 4.

Radial load was used to simulate three operation states of rolling bearings. The rotation speeds of the drive motor were adjusted to 400 rpm, 800 rpm and 1200 rpm, respectively. AE signals of rolling bearings were acquired at a sampling frequency of 96 kHz. There were 10 samples in each rolling bearing operation state. Thus, 90 samples were collected for each bearing condition and there were a total of 630 samples. The AE signals of seven rolling bearing conditions are shown in Fig. 5.

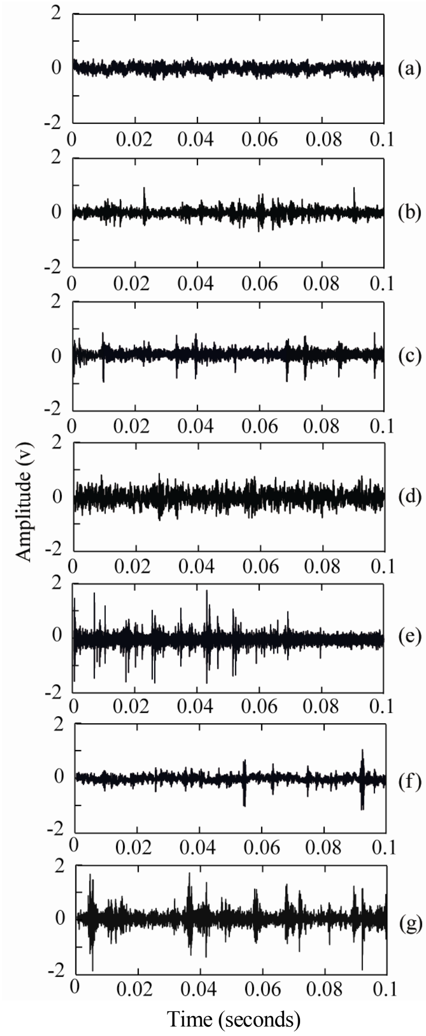


Fig. 5. AE signals of seven rolling bearing conditions: (a) NC; (b) IRF-0.4 mm; (c) IRF-0.8 mm; (d) ORF-0.4 mm; (e) ORF-0.8 mm; (f) REF; (g) CF.

4.2 Experimental results

By adding white noise, ensemble empirical mode decomposition (EEMD) can automatically project the composed components of a signal onto a uniform reference frame to solve mode mixing problem [37]. Consequently, it has been applied in extracting fault features of rolling bearings [38, 39]. In this experiment, EEMD is used to decompose the AE signals into several intrinsic mode functions (IMFs). Both time domain and frequency domain features are extracted to characterize rolling bearing conditions. These features contain standard deviation, kurtosis, shape factor and impulse factor in the time domain, and mean frequency, root mean square frequency, standard deviation frequency and spectrum peak ratio in the frequency domain. The first three maximum IMF values of each feature are regarded as fault features of rolling bearings. Accordingly, a total of 24 fault features are acquired to form condition attributes.

To validate the performance of the proposed method, 504 samples are regarded as training samples and 126 samples are used for test in this study. The ratio of training to test samples is 4:1. The attribute reduction algorithm based on MMG-DTRS is applied to acquire a lower dimensional condition attribute set from the training samples. The threshold α is set to 0.7. By using principle component analysis (PCA), the scatter plots of the training samples with the entire condition attributes and the ones with the lower dimensional condition attribute set are shown in Fig. 6. It can be observed that the lower dimensional condition attribute set provides better separation with a distinct clustering distribution among seven rolling bearing conditions. It is because the condition attributes in the acquired lower dimensional condition attribute set are

Table 1. Diagnosis accuracy of seven rolling bearing conditions.

Condition	Ratio of training to test samples	Training accuracy (%)	Test accuracy (%)	Average accuracy (%)
NC	4:1	100	77.78	95.56
IRF-0.4 mm	4:1	100	83.33	96.67
IRF-0.8 mm	4:1	100	88.89	97.78
ORF-0.4 mm	4:1	100	88.89	97.78
ORF-0.8 mm	4:1	100	94.44	98.89
REF	4:1	100	83.33	96.67
CF	4:1	100	88.89	97.78

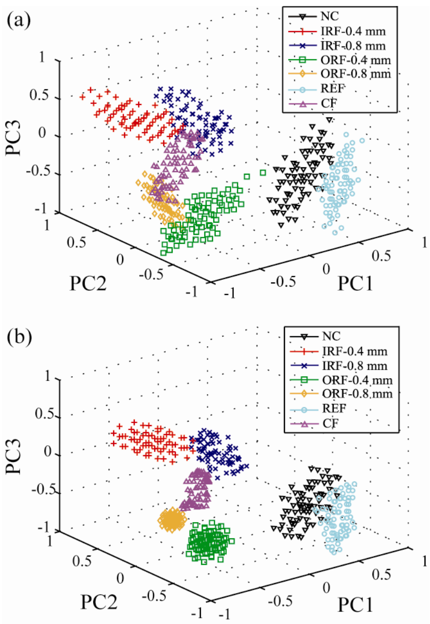


Fig. 6. Scatter plots of the training samples: (a) With the entire condition attributes; (b) with the lower dimensional condition attribute set.

more sensitive during clustering performance, which avoids the interference of irrelevant or redundant condition attributes.

NNBC is constructed to classify rolling bearing conditions in the test samples. Diagnosis accuracy of seven rolling bearing conditions is listed in Table 1, where average accuracy is calculated according to all the samples of each rolling bearing condition. As listed in Table 1, the average accuracy of each condition is more than 95 %. Thus, the proposed method can accurately identify fault categories and fault severities of rolling bearings.

4.3 Comparison analyses

Toanalyze the influence of the threshold α on average accuracy, the attribute reduction algorithm based on MMG-DTRS and the other one based on DTRS are, respectively, applied to acquire lower dimensional condition attribute sets. Then, rolling bearing conditions in test samples are classified by using NNBC. Relation curves between average accuracy and the threshold α are depicted in Fig. 7. It can be observed that average accuracy increases first and then decreases along with the increase of the threshold α . It is because smaller threshold values will lead to constraint conditions being loose, while irrelevant or redundant condition attributes are probably retained. Larger threshold values will lead constraint conditions to be harsh, while sensitive and essential condition attributes are probably removed. Accordingly, the best diagnostic performance can be achieved by reasonable threshold value selection in practical applications. Although average accuracy is influenced by the threshold α , the proposed method still possesses higher average accuracy than the other one based on DTRS. The reason is that MMG-DTRS can realize data analysis from multi-level view and describe a target concept through fusing multiple granular structures rather than a single one. Therefore, the proposed method is exactly suitable to identify various types of rolling bearing faults.

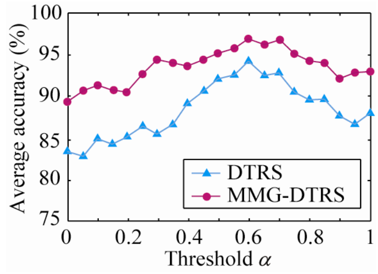


Fig. 7. Relation curves between average accuracy and the threshold α .

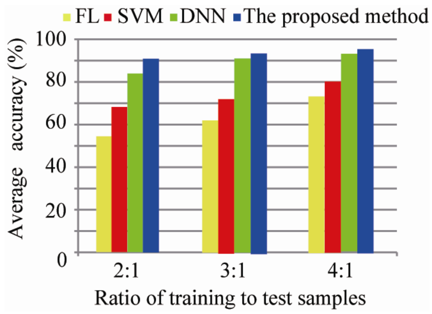


Fig. 8. Diagnosis results of four methods.

To further validate the effectiveness, the proposed method is compared to the traditional data-driven fault diagnosis methods (FL, SVM and DNN) using the same samples. In this experiment, fuzzy toolbox available in MATLAB 2016a is used to construct the fuzzy inference engine. Fault diagnosis based on SVM is performed by MATLAB SVM toolbox, whose penalty parameter C and Gaussian kernel parameter γ are set to 120 and 10, respectively. A three-layer DNN is applied and the number of hidden layer nodes is set as 100. The ratios of training to test samples are 2:1, 3:1 and 4:1, respectively. Fig. 8 illustrates diagnosis results of four methods. It can be noted that the average accuracy of four methods increases along with the ratio enhancement. The higher the ratio of training to test samples is, the better the diagnosis effectiveness is. Compared with the other methods, the proposed method always achieves the highest average accuracy with three ratios. This is because more training samples possess more diagnosis information, which can realize more accurate determination of the model structure and parameters. Besides, the lower dimensional condition attribute set provides better separation with a distinct clustering distribution among rolling bearing conditions. In addition, NNBC ignores the independence assumption among attributes and replaces marginal PDF estimation by joint PDF estimation. We suggest using the proposed method under the following condition. A great number of vibration signals can be acquired from historical records. Moreover, these historical records contain various conditions of rolling bearings.

4.4 Discussions

(1) According to the above experimental results, the diagnosis accuracy of the proposed method obviously outperforms the traditional data-driven fault diagnosis methods. The main reasons are summarized as follows. The attribute reduction algorithm based on MMG-DTRS is applied to acquire sensitive condition attributes, which can maximize classification separation and avoid the interference of irrelevant or redundant condition attributes. Also, NNBC based on the optimal bandwidth selection ignores the independence assumption among attributes and replaces marginal PDF estimation by joint PDF estimation.

(2) In the comparison analyses, the proposed method obtains higher average accuracy than the other method based on DTRS, despite the influence of the threshold α. The reasons are that MMG-DTRS retains fault-tolerant ability of DTRS, realizes data analysis from multi-level view, and overcomes the shortcoming that constraint conditions of pessimistic MGRS model are excessively harsh. In addition, the attribute reduction algorithm based on MMG-DTRS can remove redundant condition attributes and decrease the number of random variables in NNBC. Thus, the proposed method is capable of improving diagnosis accuracy and reducing computational burden.

(3) In the experimental investigations, seven rolling bearing conditions covering typical fault categories and different fault severities are used to validate the performance of the proposed method. Accordingly, it is a representative case of rolling bearing fault diagnosis. According to the comparison analyses, the comparison results demonstrate the effectiveness, reliability and generalization ability of the proposed method. Thus, this method is exactly suitable to identify various types of rolling bearing faults.

5. Conclusions

A novel fault diagnosis method of rolling bearings combining MMG-DTRS with NNBC was put forward. MGRSs, MMG-DTRS, and NNBC were first reviewed. Then, the fault diagnosis framework was illustrated. A fault diagnosis experiment of rolling bearings was utilized to validate the performance of this method. The comparison analyses demonstrate that the diagnosis accuracy of this method is much superior to that of the traditional methods. Besides, the proposed method is capable of improving diagnosis accuracy and reducing computational burden. Therefore, it is effective and reliable in identifying fault categories and fault severities of rolling bearings.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (61673142, and 51275136), and the University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province (2016034). The authors are also grateful to the editors and the anonymous reviewers for their helpful comments and constructive suggestions.

References

[1] H. D. M. de Azevedo, A. M. Araújo and N. Bouchonneau, A review of wind turbine bearing condition monitoring: State of the art and challenges, *Renewable and Sustainable Energy Reviews*, 56 (2016) 368-379.

[2] F. T. Wang, C. X. Liu, W. S. Su, Z. G. Xue, H. K. Li and Q. K. Han, Condition monitoring and fault diagnosis methods for low-speed and heavy-load slewing bearings: A literature review, *J. of Vibroengineering*, 19 (5) (2017) 3429-3444.

[3] Y. Gui, Q. K. Han and F. L. Chu, A vibration model for fault diagnosis of planetary gearboxes with localized planet bearing defects, *J. of Mechanical Science and Technology*, 30 (9) (2016) 4109-4119.

[4] Y. Zhou, L. Li, D. Wang, M. He and D. He, A new method to classify railway vehicle axle fatigue crack AE signal, *Applied Acoustics*, 131 (2018) 174-185.

[5] N. Upadhyay and P. K. Kankar, Diagnosis of bearing defects using tunable Q-wavelet transform, *J. of Mechanical Science and Technology*, 32 (2) (2018) 549-558.

[6] M. Cerrada, R. V. Sánchez, C. Li, F. Pacheco, D. Cabrera, J. V. de Oliveira and R. E. Vasquez, A review on data-driven fault severity assessment in rolling bearings, *Mechanical Systems and Signal Processing*, 99 (2018) 169-196.

[7] W. Deng, H. M. Zhao, L. Zou, G. Y. Li, X. H. Yang and D. Q. Wu, A novel collaborative optimization algorithm in solving complex optimization problems, *Soft Computing*, 21 (15) (2017) 4387-4398.

[8] W. Deng, H. M. Zhao, X. H. Yang, J. X. Xiong, M. Sun and B. Li, Study on an improved adaptive PSO algorithm for solving multi-objective gate assignment, *Applied Soft Computing*, 59 (2017) 288-302.

[9] M. Gan, C. Wang and C. A. Zhu, Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings, *Mechanical Systems and Signal Processing*, 72-73 (2016) 92-104.

[10] X. J. Guo, L. Chen and C. Q. Shen, Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis, *Measurement*, 93 (2016) 490-502.

[11] H. D. Shao, H. K. Jiang, Y. Lin, X. Q. Li, A. Moosavian and M. Khazaee, A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders, *Mechanical Systems and Signal Processing*, 102 (2018) 278-297.

[12] H. D. Shao, H. K. Jiang, H. Z. Zhang and T. C. Liang, Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network, *IEEE Transactions on Industrial Electronics*, 65 (3) (2018) 2727-2736.

[13] W. Zhang, C. H. Li, G. L. Peng, Y. H. Chen and Z. J. Zhang, A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load, *Mechanical Systems and Signal Processing*, 100 (2018) 439-453.

[14] Y. J. Li, W. H. Zhang, Q. Xiong, D. B. Luo, G. M. Mei and T. Zhang, A rolling bearing fault diagnosis strategy based on improved multiscale permutation entropy and least squares SVM, *J. of Mechanical Science and Technology*, 31 (6) (2017) 2711-2722.

[15] Y. B. Li, M. Q. Xu, H. Y. Zhao and W. H. Huang, Hierarchical fuzzy entropy and improved support vector machine based binary tree approach for rolling bearing fault diagnosis, *Mechanism and Machine Theory*, 98 (2016) 114-132.

[16] R. N. Liu, B. Y. Yang, X. L. Zhang, S. B. Wang and X. F. Chen, Time-frequency atoms-driven support vector machine method for bearings incipient fault diagnosis, *Mechanical Systems and Signal Processing*, 75 (2016) 345-370.

[17] Y. B. Li, M. Q. Xu, Y. Wei and W. H. Huang, A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree, *Measurement*, 77 (2016) 80-94.

[18] R. Ziani, A. Felkaoui and R. Zegadi, Bearing fault diagnosis using multiclass support vector machines with binary particle swarm optimization and regularized Fisher's criterion, *J. of Intelligent Manufacturing*, 28 (2) (2017) 405-417.

[19] J. D. Zheng, H. Y. Pan and J. S. Cheng, Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines, *Mechanical Systems and Signal Processing*, 85 (2017) 746-759.

[20] M. Straczkiewicz, P. Czop and T. Barszcz, The use of a fuzzy logic approach for integration of vibration-based diagnostic features of rolling element bearings, *J. of Vibroengineering*, 17 (4) (2015) 1760-1768.

[21] R. Ziani, A. Felkaoui and R. Zegadi, Bearing fault diagnosis based on multi-scale permutation entropy and adaptive neuro fuzzy classifier, *J. of Vibration and Control*, 21 (3) (2015) 461-467.

[22] Z. B. Xu, Y. R. Li, Z. G. Wang and J. P. Xuan, A selective fuzzy ARTMAP ensemble and its application to the fault diagnosis of rolling element bearing, *Neurocomputing*, 182 (2016) 25-35.

[23] X. B. Sun, J. W. Tan, Y. Wen and C. S. Feng, Rolling bearing fault diagnosis method based on data-driven random fuzzy evidence acquisition and Dempster-Shafer evidence theory, *Advances in Mechanical Engineering*, 8 (1) (2016) 1687814015624834.

[24] C. Li, L. Ledo, M. Delgado, M. Cerrada, F. Pacheco, D. Cabrera, R. V. Sanchez and J. V. de Oliveiraae, A Bayesian approach to consequent parameter estimation in probabilistic fuzzy systems and its application to bearing fault classification, *Knowledge-Based Systems*, 129 (2017) 39-60.

[25] W. Deng, S. J. Zhang, H. M. Zhao and X. H. Yang, A novel fault diagnosis method based on integrating empirical wavelet transform and fuzzy entropy for motor bearing, *IEEE Access*, Doi: 10.1109/ACCESS.2018.2834540.

[26] Y. H. Qian, J. Y. Liang, Y. Y. Yao and C. Y. Dang, MGRS: A multi-granulation rough set, *Information Sciences*, 180 (6) (2010) 949-970.

[27] Y. H. Qian, J. Y. Liang and C. Y. Dang, Incomplete multigranulation rough set, *IEEE Transactions on Systems, Man, and Cybernetics-Part A: System and Humans*, 40 (2) (2010) 420-431.

[28] Y. H. Qian, S. Y. Li, J. Y. Liang, Z. Z. Shi and F. Wang, Pessimistic rough set based decisions: A multigranulation fusion strategy, *Information Sciences*, 264 (2014) 196-210.

[29] A. H. Tan, W. Z. Wu and Y. Z. Tao, On the belief structures and reductions of multigranulation spaces with decisions, *International J. of Approximate Reasoning*, 88 (2017) 39-52.

[30] C. Zhang, Y. H. Zhai, D. Y. Li and Y. M. Mu, Steam turbine fault diagnosis based on single-valued neutrosophic multigranulation rough sets over two universes, *J. of Intelligent and Fuzzy Systems*, 31 (6) (2016) 2829-2837.

[31] C. Zhang, D. Y. Li, Y. M. Mu and D. Song, An interval-valued hesitant fuzzy multigranulation rough set over two universes model for steam turbine fault diagnosis, *Applied Mathematical Modelling*, 42 (2017) 693-704.

[32] Y. Y. Yao, S. K. M. Wong and P. Lingras, A decision-theoretic rough set model, *Methodologies for intelligent systems 5*, North-Holland, New York, USA (1990) 17-24.

[33] Y. H. Qian, H. Zhang, Y. L. Sang and J. Y. Liang, Multigranulation decision-theoretic rough sets, *International J. of Approximate Reasoning*, 55 (1) (2014) 225-237.

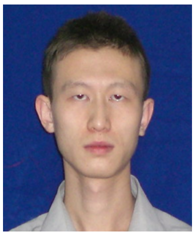
[34] Y. L. He, R. Wang, S. Kwong and X. Z. Wang, Bayesian classifiers based on probability density estimation and their applications to simultaneous fault diagnosis, *Information Sciences*, 259 (2014) 252-268.

[35] M. Y. Asr, M. M. Ettefagh, R. Hassannejad and S. N. Razavi, Diagnosis of combined faults in rotary machinery by non-naive Bayesian approach, *Mechanical Systems and Signal Processing*, 85 (2017) 56-70.

[36] X. Z. Wang, Y. L. He and D. D. Wang, Non-naive Bayesian classifiers for classification problems with continuous attributes, *IEEE Transactions on Cybernetics*, 44 (1) (2014) 21-39.

[37] Z. H. Wu and N. E. Huang, Ensemble empirical mode decomposition: A noise assisted data analysis method, *Ad-*

vances in Adaptive Data Analysis, 1 (1) (2009) 1-41.
[38] Y. G. Lei, Z. J. He and Y. Y. Zi, EEMD method and WNN for fault diagnosis of locomotive roller bearings, *Expert Systems with Applications*, 38 (6) (2011) 7334-7341.
[39] Y. Imaouchen, M. Kedadouche, R. Alkama and M. Thomas, A frequency-weighted energy operator and complementary ensemble empirical mode decomposition for bearing fault detection, *Mechanical Systems and Signal Processing*, 82 (1) (2017) 103-116.



Jun Yu received his M.S. and Ph.D. in Mechanical and Electrical Engineering from Harbin Institute of Technology, in China, in 2009 and 2017, respectively. He has been working in the School of Harbin University of Science and Technology. His main research interests include mechanical system fault diagnosis, knowledge discovery and data mining.