World Scientific Research Journal (WSRJ) ISSN: 2472-3703
www.wsr-j.org

Hybrid filter and wrapper feature selection scheme based on ReliefF and NSGA-II for gear fault diagnosis

Runze Shi[1], Bing Li[2], Junning Zhang[3], Xu Li[4], Yuliang Lu[5]

[1] Forth department, The Army Engineering University of PLA, No.97, He-ping West Road, Shi Jia-zhuang, 050003, He Bei province, PR China.

[2] Forth department, The Army Engineering University of PLA, No.97, He-ping West Road, Shi Jia-zhuang, 050003, He Bei province, PR China.

[3] Forth department, The Army Engineering University of PLA, No.97, He-ping West Road, Shi Jia-zhuang, 050003, He Bei province, PR China.

[4,5] 5413 Military Representative Office, Shi Jia-zhuang, 050003, He Bei province, PR China.

shirunze1993@163.com, rommandy@163.com, 18632144905m@sina.cn

Abstract: In this paper, we present a hybrid filter and wrapper (HFW) feature selection scheme based on ReliefF algorithm and non-dominated sorting genetic algorithms II (NSGA-II) for gear fault diagnosis. We conduct an experiment on a gearbox, in which four different gear operating states including one normal state and three fault states were simulated, to evaluate the presented HFW scheme. The original features are calculated based on EMD, Autoregressive (AR) model, statistical techniques, singular value decomposition and entropy theories. Three different classifiers are adopted to evaluate and compare the proposed HFS scheme with the separate filter and wrapper feature selection methods. Experimental results have demonstrated that the proposed HFW feature selection scheme gained higher classification accuracies with lower dimension feature subsets than the separate filter and wrapper feature selection approaches. A very satisfying performance can be achieved by the presented HFW feature selection scheme based on ReliefF and NSGA-II for gear fault diagnosis.

Keywords: Feature selection, Hybrid filter and wrapper, ReliefF, non-dominated sorting genetic algorithms II (NSGA-II), gear, fault diagnosis.

---

1. INTRODUCTION

Gear is one of the most widely used mechanical components in modern industrial application areas. Unexpected failure of the gear may cause significant causalities and economic losses. For this reason, the need to increase reliability against possible gear failures has aroused great interest in recent years. Vibration signal analysis is one of the most effective techniques for detecting gear faults due to its effectiveness and ease to measure [1-4]. Although often the visual inspection of the frequency domain features of the measured signals is adequate to identify the faults, most of the techniques available require a good deal of expertise to apply them successfully. Consequently, various intelligent techniques such as artificial neural networks (ANN), support vector machine (SVM), fuzzy logic and evolving algorithms (EA) have been successfully applied to automated detection and diagnosis of machine conditions over the past few years [5-12].

For any intelligent fault diagnosis systems, feature extraction and feature selection schemes can be regarded as the most two important steps. Feature extraction is a mapping process from the measured signal space to the feature space. Representative features associated with the conditions of machinery components should be extracted by using appropriate signal processing and calculating approaches. Usually, we are inclined to extract features as many as possible to characterize the vibration signals completely. However, it must be noted that there are many irrelevant and redundant features existed in the original feature set. Employing the whole feature set will lead to the large cost of time and lower performance for fault diagnosis. Therefore, in order to improve the diagnosis accuracy and reduce the cost of computation, the most sensitive parameters to characterize the machine conditions should be selected from the original feature set.

Filter and wrapper methods can be regarded as the two mainly categories of feature selection approaches in literatures. Several researches has been done by utilizing the filter method or wrapper method for mechanical fault diagnosis utilizing Fisher's criterion [13], distance evaluation technique [14, 15], decision tree [16, 17] and evolution algorithm (EA) combined with ANN and SVM [5, 18-22]. They have largely improved the efficiency and accuracy of mechanical fault diagnosis in practice. Despite the successive applications of filter or wrapper methods in fault diagnosis, it should be noted that the wrapper and filter methods can complement each other. The filter methods can search through the feature space efficiently but usually fail to obtain a good accuracy, while the wrappers can provide good accuracy but require much computation time. Therefore, it is very desirable to combine the filter and wrapper methods to achieve high efficiency and accuracy simultaneously [23-26].

In this work, we present a hybrid feature selection scheme combing filter and wrapper methods base on the ReliefF [27] and the improved non-dominated sort genetic algorithm (NSGA-II) [28] for gear fault diagnosis. Experiments were conducted on a single stage gearbox. Four different gear conditions including normal state and three fault states were simulated in the experiments. Vibration signals collected from the gearbox were employed for identifying the gear faults. The empirical mode decomposition (EMD) technique is used to process the

vibration signal and the AR model, statistical techniques, singular value decomposition (SVD) and entropy theory are employed to calculate the original features. Then the presented HFW feature selection scheme is applied to find an optimal feature subset for gear fault diagnosis. We evaluate and compare the performance of the proposed HFW scheme on the gear dataset with the separate filter and wrapper approaches. Experimental results have indicated the superiority of the proposed HFW method on computation cost and classification accuracies in gear fault diagnosis.

The remainder of this work is organized as following: Section 2 describes the feature extraction method based on the EMD, AR model, statistical techniques, SVD and entropy methods. In Section 3, the hybrid filter and wrapper feature selection scheme based on ReliefF and NSGA-II is detailed. Section 4 presents the application results of the proposed HFW feature selection scheme for gear fault diagnosis. The conclusions of this paper are summarized in Section 5.

2. FEATURE EXTRACTION BASED ON EMD

2.1 EMD method

The empirical mode decomposition (EMD) technique, which is proposed by Huang in 1998 [29], has been applied widely in the field of mechanical fault detection in the past few years [15, 30-34]. The main idea of EMD is decomposing a non-stationary signal into some intrinsic mode functions (IMF), which represent the simple oscillatory mode imbedded in the signal. The detail description and implementation of the EMD method can be found in reference [29]. At the end of the EMD process, a signal $x(t)$ is decomposed into a collection of IMFs c_i and a residue r_M, which satisfy the following equation:

$$x(t) = \sum_{i=1}^{M} c_i + r_M \tag{1}$$

The IMFs $c_1, c_2, \cdots, c_M$ include different frequency bands ranging from high to low. The frequency components contained in each frequency band are different and they change with the variation of signal $x(t)$.

2.2 Feature extraction

2.2.1 Statistical parameters.

In this paper, ten time domain parameters including the maximum, minimum, peak-peak, RMS, kurtosis, skewness, crest factor, impulse factor, shape factor and clearance factor are selected as the statistical parameters for characterizing the signal. The detail definitions of these parameters are summarized in Table 1.

Table1. Definition of the ten statistical parameters

$$p_0 = \sum_{n=1}^{N} x(n)/N$$

$$p_1 = \max(|x(n)|)$$

$$p_2 = \min(|x(n)|)$$

$$p_3 = p_1 - p_2$$

$$p_4 = \sqrt{\sum_{n=1}^{N} x(n)^2 / N}$$

$$p_5 = \frac{\sum_{n=1}^{N} (x(n) - p_0)^3}{p_4^3 N}$$

$$p_6 = \frac{\sum_{n=1}^{N} (x(n) - p_0)^4}{p_4^4 N}$$

$$p_7 = p_1 / p_4$$

$$p_8 = \frac{p_1}{\sum_{n=1}^{N} |x(n)| / N}$$

$$p_9 = \frac{p_4}{p_0}$$

$$p_{10} = \frac{p_1}{(\frac{1}{N}\sum_{n-1}^{N} \sqrt{|x(n)|})^2}$$

where $x(n)$ is a signal series, $n = 1,2,\cdots,N$.

2.2.2 AR model coefficients.

Autoregressive (AR) model is a time sequence analysis method whose parameters comprise important information of the system condition. Therefore, the parameters of the AR model can be effectively used to analyze the condition variation of signals [35].

For each IMF $c_i(t)$, the following AR model can be established:

$$c_i(t) + \sum_{k=1}^{m} \varphi_{ik} c_i(t-k) = e_i(t) \quad i = 1,2,\cdots,M \tag{2}$$

where m is the order of the AR model and $\varphi_{ik}(k = 1,2,\cdots m)$ are the coefficients of the AR model of $c_i(t)$; $e_i(t)$ is the remnant of the model. In our work, the feature vector $V_i = [\varphi_{i1}, \varphi_{i2}, \cdots, \varphi_{im}]$ is selected as features to characterize the IMF $c_i(t)$ ($i = 1,2,\cdots,M$).

2.2.3 Singular value and singular value entropy of IMFs.

The M IMFs $c_i(t)$ ($i = 1,2,\cdots,M, t = 1,2,\cdots,N$) can be regarded as a $M \times N$ dimensional matrix R. According to the singular value decomposition theory, a $M \times N$ dimensional matrix R can be decomposed to a $M \times L$ dimensional matrix U, a $L \times L$ dimensional diagonal matrix Λ and a $L \times N$ dimensional matrix V, which can be represented as

$$R_{M\times N} = U_{M\times L} \Lambda_{L\times L} V_{L\times N} \tag{3}$$

The singular values λ_i ($i = 1,2,\cdots,L$) lie on the diagonal of Λ and be sorted in a descend order, means that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L \geq 0$. Let p_S denotes the singular value distribution and can be obtained by

$$p_S(i) = \lambda_i \ / \ \sum_{i=1}^{L} \lambda_i \tag{4}$$

According to the defined singular value distribution p_S, the IMFs singular entropy I_{SE} can be defined as

$$I_{SE} = -\sum_{i=1}^{L} p_S(i) \log p_S(i) \tag{5}$$

3. HYBRID FILTER AND WRAPPER FEATURE SELECTION SCHEME BASED ON RELIEFF AND NSGA-II

3.1 Hybrid filter and wrapper feature selection scheme

Filters and wrappers are the two mainly categories of feature selection algorithms in literatures. Filter methods evaluate the goodness of the feature subset by using the intrinsic characteristic of the data. They are relatively computationally cheap since they do not involve the induction algorithm. However, they also take the risk of selecting subsets of features that may not match the chosen induction algorithm. Wrapper methods, on the contrary, directly use the classifiers to evaluate the feature subsets. They generally outperform filter methods in terms of prediction accuracy, but they are generally computationally more intensive [36-39]. In summary, wrapper and filter methods can complement each other, in that filter methods can search through the feature space efficiently while the wrappers can provide good accuracy. It is desirable to combine the filter and wrapper methods to achieve high efficiency and accuracy simultaneously.

In this work, we present a hybrid feature selection combing filter and wrapper feature selection technique base on the ReliefF [27] and the improved non-dominated sort genetic algorithm (NSGA-II) [28]. In our presented hybrid filter and wrapper (HFW) feature selection scheme, there are two steps involved. In the first stage, a candidate feature subset is chosen according to the relevance values calculated by ReliefF from the original feature set. Then at the second stage, classifier combined with NSGA-II is adopted to find a more compact feature subset from the candidate feature subset. In this stage, feature selection problem is defined as a multi-objective problem dealing with two competing objectives, mean lesser features and lower classification error rate. Fig. 1 displays the flowchart of the presented HFW feature selection scheme.

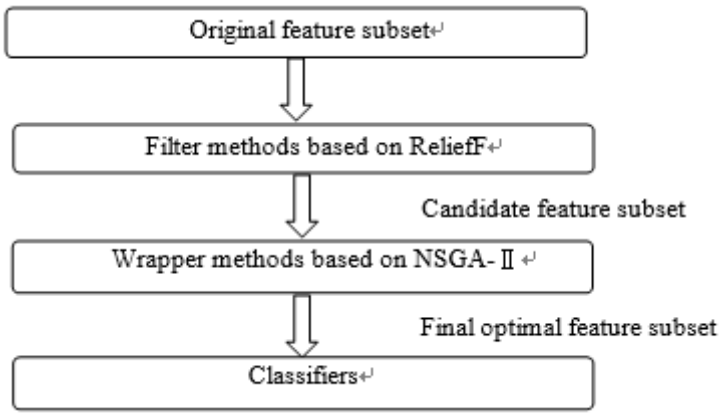

Fig. 1. Flowchart of the presented HFW feature selection scheme

3.2 Filter method based on ReliefF method

ReliefF is an extension of the popular Relief algorithm, which are actually general feature estimators and have been used successfully as attribute weighting method [27]. The key idea of the original Relief algorithm is to estimate the quality of attributes according to how well their values distinguish between instances that are near to each other. The original Relief cannot deal with incomplete data and is limited to two-class problems. As an extension of the Relief algorithm, ReliefF can handle multiple classes and is more robust to noise.

Similarly to Relief, ReliefF randomly selects an instance Ri, but then searches for k of its nearest neighbors from the same class, called nearest hits Hj, and also k nearest neighbors from each of the different classes, called nearest misses Mj (C). It updates the quality estimation W[A] for all attributes A_i ($i = 1, 2, \cdots, a$, a is the number of features) depending on their values for Ri , hits Hj and misses Mj (C). The whole process is repeated for m times, where m is a user-defined parameter.

The pseudo code of ReliefF is given in Table 2.

Table 2 The pseudo code of ReliefF algorithm

Algorithm ReliefF

Input: for each training instance a vector of attribute values and the class value

Output: the vector *W* of estimations of the qualities of attributes

1. set all weights $W[A] := 0$;
2. for *i* := 1 to *m* do begin
3. randomly select an instance R_i ;
4. find *k* nearest hits H_j ;
5. for each class $C \neq class(R_i)$ do
6. from class *C* find *k* nearest misses M_j (*C*);
7. end
9. for *A* := 1 to a do
10. $$W[A] := W[A] - \sum_{j=1}^{k} diff(A, R_i, H_j)/(m \cdot k) + \sum_{C \neq class(R_i)} [\frac{P(C)}{1 - P(class(R_i))} \sum_{j=1}^{k} diff(A, R_i, M_j(C))]/(m \cdot k)$$
11. end

12 end

3.3 Wrapper method based on NSGA-II

3.3.1 A brief review on NSGA-II.

The presence of multiple objectives in practical problems has given rise to the rapid development of multi-objective evolutionary algorithms over the past few years. Non-dominated Sorting Genetic Algorithm (NSGA), which was suggested by Goldberg and implemented by Srinivas and Deb[40], has been proved to be an effective approach for

multi-objective optimization problems. However, NSGA is still suffering three main drawbacks, mean the high computational complexity of non-dominated sorting, lack of elitism and requirement for specifying the sharing parameter.

As an improved version of the NSGA, NSGA-II was introduced by Deb in 2002 [28]. The NSGA-II overcame the original NSGA defects by introducing the fast non-dominated sorting algorithm to alleviate computational complexity, the elitist-preserving mechanism to speed up the evolution and the crowded comparison operator to avoid specifying the sharing parameter. It has been verified that the NSGA-II is able to maintain a better spread of solutions and converge better in the obtained non-dominated front compared to other similar elitist multiple objectives evolution algorithms (MOEAs). More details about the description and implementation can be found in reference [28].

3.3.2 Wrapper feature selection using NSGA-II.

In most cases of conventional wrapper methods for fault diagnosis, the feature selection problem was formulized as a single objective problem [5, 18, 20, 22, 41]. However, the feature selection is inherently a multi-objective problem, which deals with two competing objectives mean the feature dimension and the classification accuracy. An optimal feature set has to be of a minimal number of features and has to produce the minimum classification error rate.

In this work, we formulate the feature selection problem for gear fault diagnosis to be a multi-objective problem. The NSGA-II mentioned above is utilized to optimize the two objectives, mean the minimal number of features and minimum classification error rate. A step-by-step procedure for solving the feature selection problem by utilizing the NSGA-II is illustrated in Fig. 2.

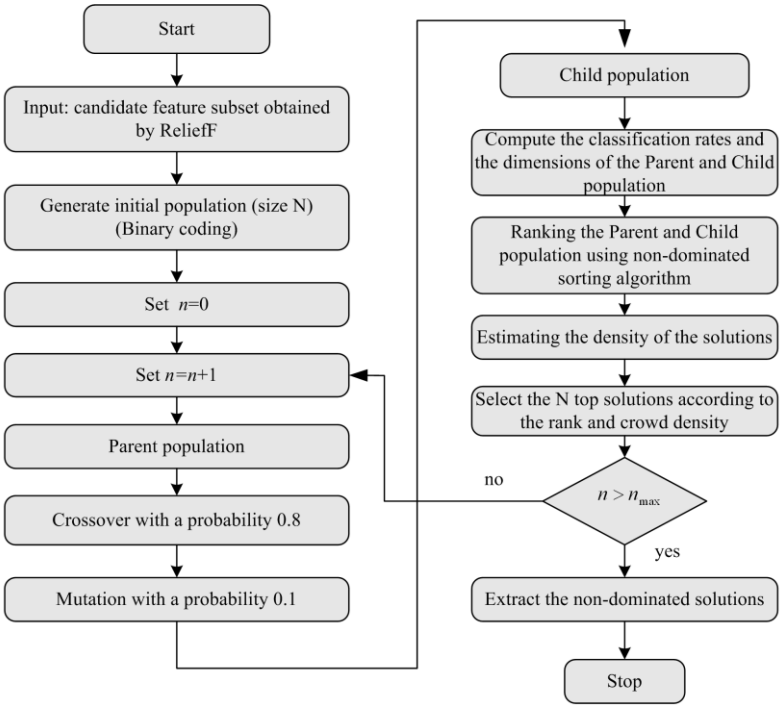

Fig. 2. The schematic for applying the NSGA-Ⅱ to the feature selection

3.3.3 Implementation issues for wrapper feature selection using NSGA-II.

For wrapper feature selection approach, there are several factors for controlling the process of NSGA-II while searching the sub-optimal feature subsets for classifiers. To apply NSGA-II to feature selection, we focus on the following issues.

(1) Fitness functions

Two competing objectives are defined as the fitness functions: the first was minimization of the number of used features and the second was minimization of the classification error rate. Three popular classifiers means the K nearest neighbor classifier (KNNC) [42] , Naïve Bayes classifier) (NBC) [43] and least-square support vector machine (LS-SVM) [44] were employed as induction algorithms to implement and evaluate the proposed feature selection approach. The KNNC and NMC were implemented by utilizing the Matlab Toolbox for Pattern Recognition (PRTools 4.1) [45]. The LS-SVM was implemented by the LS-SVMlab1.5, which can be downloaded from [46].

(2) Encoding Scheme

The binary coding system was used to represent the chromosome in this investigation. For chromosome representing the feature subsets, the bit with value ‘1’ represents the feature is selected, and ‘0’ indicates feature is not selected, as shown in Fig. 3.

$$\underbrace{\boxed{0}\boxed{1}\boxed{0}\boxed{0}\boxed{1}\boxed{\ldots}\boxed{0}\boxed{1}\boxed{0}}_{n}$$

Fig. 3. Representation of chromosome as a binary bit string for feature selection

(3) Genetic operators

Genetic operator consists of two basic operators, i.e., crossover and mutation. The used crossover technique was the uniform crossover consisting on replacing genetic material of the two selected parents uniformly in several points. The mutation operator used in this work was implemented as conventional mutation operator operating on each bit separately and changing randomly its value.

4. RESULTS AND DISCUSSION

4.1 Experimental system of gear

The vibration data of gear used in this paper was acquired from a single stage gearbox. Fig. 4 displays the diagram of the experimental system used in this work. The experimental system includes a single stage gearbox, an ac motor and a magnetic brake. The ac motor is used to drive the gearbox and the rotating speed is controlled by a speed controller, which allows the tested gearbox to operate under various speeds. The load is provided by the magnetic brake connected to the output shaft and the torque can be adjusted by a brake controller. Four gear states, including normal, gear tooth wear, gear root crack and gear tooth broken, were simulated in the experiment. The vibration signals were picked up by accelerator sensors attached on the base of the gearbox. The sample frequency was set to be 2048 Hz and the

sample length was set to be 2048 points. Fig. 5 demonstrates the waveforms of the four gear operating states in time domain.

For each of the four operating condition, 72 samples were collected and the whole data set consists of 288 samples. Then the dataset is split into two sets: 144 samples for training and 144 samples for testing.

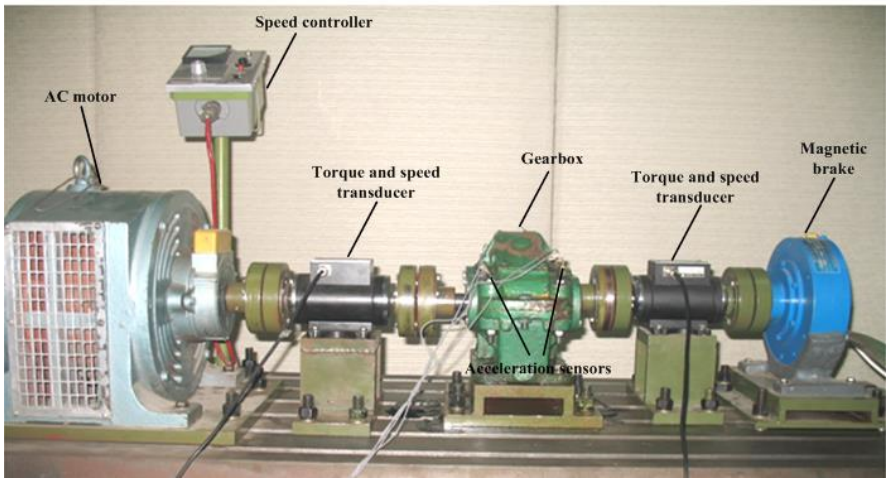

Fig. 4. The experimental gearbox

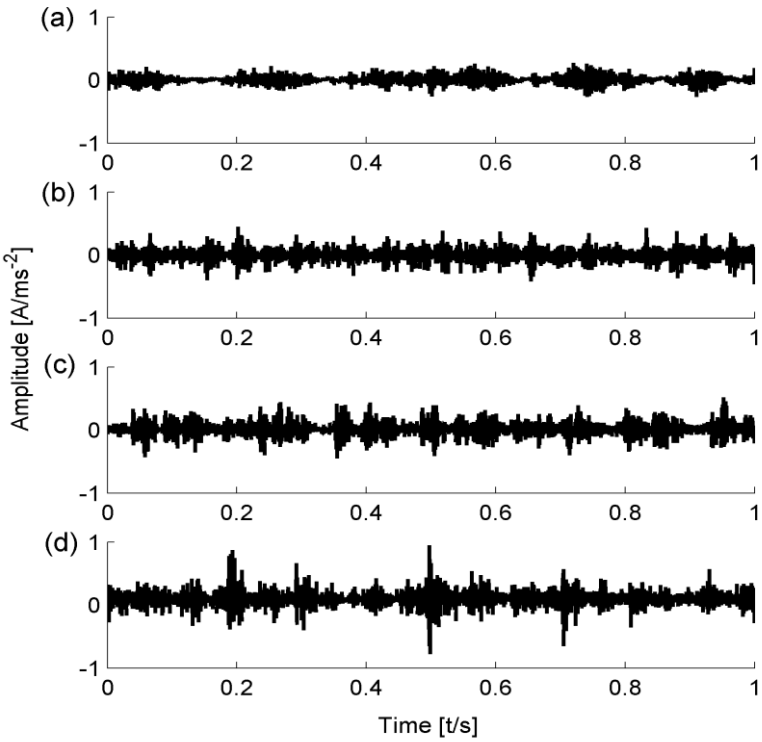

Fig. 5. Vibration signals from four states of gear:
(a) normal; (b) gear tooth wear; (c) gear root crack; (d) gear tooth broken

4.2 Feature extraction based on EMD

To extract more information, each of these raw signals was decomposed via the EMD method. In this paper, the first six IMFs, which containing almost all the validation information of

original signal, were selected for further analysis. Fig. 6 illustrates the first six IMFs of the vibration signal from gear with normal state.

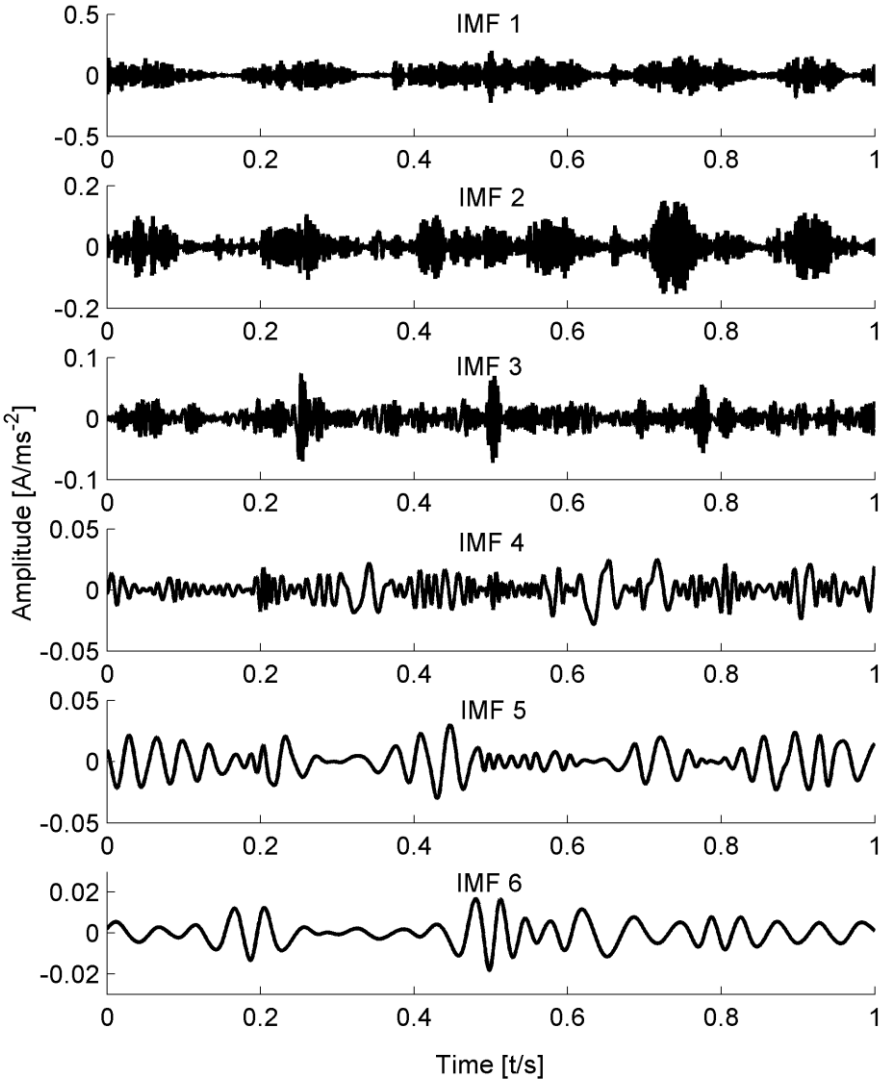

Fig. 6. The first six IMFs of vibration signal from gear with normal state decomposed by EMD

Foe each IMF decomposed by EMD, we calculate ten statistical parameters listed in Table 1. The statistical parameters of the original signal are also calculated. In this way, 70 statistical parameters can be obtained for each signal.

Then, the Autoregressive (AR) model is used to characterize the IMFs. The order of AR model is set to be 5 in this work. And the coefficients of the AR model are selected as features to characterize each IMF. Consequently, we can get 30 features for each signal.

Moreover, we compute the singular values of the IMF matrix and the singular entropy, as mentioned in section 2.2.3. We can obtain 7 features for each signal by using the SVD theory.

Finally, for each sample, 70+30+7=107 features can be calculated according to the feature extraction methods presented in section 2.2. These features constitute the original feature set $F_{Original}$ for gear fault diagnosis.

4.3 Feature evaluation by ReliefF method

The training set was used to evaluate the relevance of extracted 107 features by ReliefF method. In this study, we choose three nearest neighbors when searching for the hits and misses in ReliefF. The iteration parameter m is set to be 144. Fig. 7 illustrates the relevance values of the 107 features calculated by ReliefF.

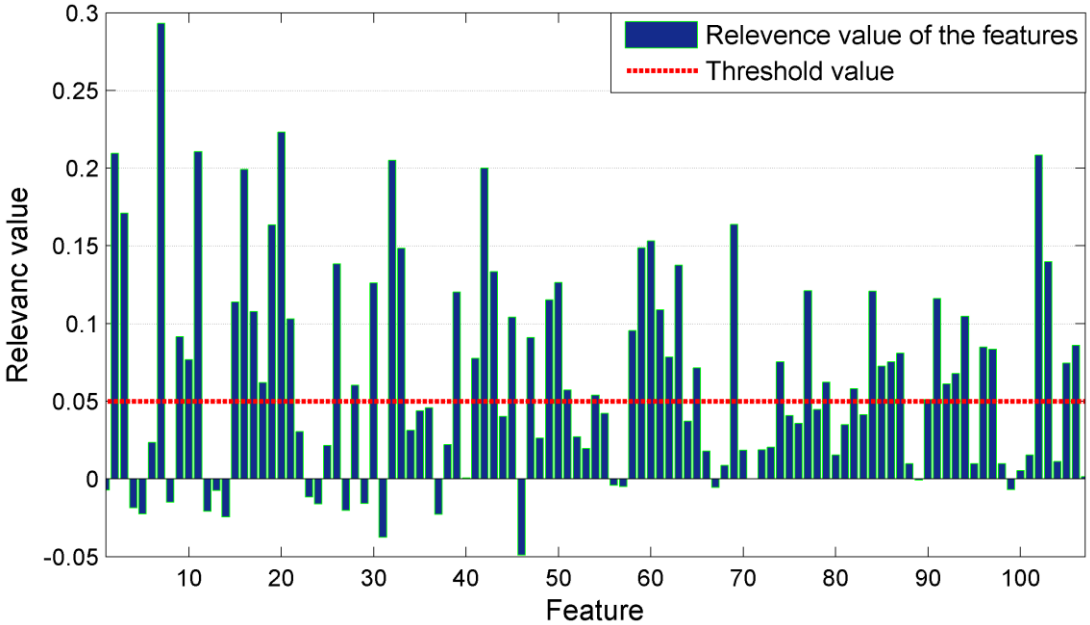

Fig. 7. Relevance values of the extracted features using ReliefF

It can be found that the relevance value of most features is larger than 0 and some is smaller than 0. The larger the value is, the more discriminative information the feature contains. A threshold 0.05 is set to select the superiority features, as shown in Fig. 7. The features that the relevance values thereof exceed the threshold,

about 59 features in our case, are selected to constitute a new feature subset. We denote this new feature subset as $F_{ReliefF}$. It will be used as a candidate feature subset for further selection by wrapper method in next subsection.

4.4 Wrapper feature selection using NSGA-II method

In this subsection, the wrapper method based on NSGA-II is used to find a more compact feature subset based on the feature subset $F_{ReliefF}$. We also conduct an experiment on applying the wrapper method directly on the original feature subset $F_{Original}$for a comparison with the two stage feature selection scheme.

Three classifiers, mean KNNC, NBC and LS-SVM as mentioned in section 3.3, are employed to evaluate the presented two-stage feature selection scheme for gear fault diagnosis. We just select the KNNC classifier to illustrate the wrapper method based on NSGA-II algorithm. For every chromosome created by NSGA-II, the dimension of features selected by this chromosome and the classification error rate based on KNNC classifier are regarded as the fitness functions. As mentioned in section 4.1, there are 288 samples were collected in total. These samples were segmented into two parts, 144 samples for training and 144 samples for testing, to calculate the classification error rate.

Other parameters of NSGA-II for feature selection are summarized in Table 3.

Table 3 Parameter settings of NSGA-II for wrapper feature selection

Parameters of NSGA- II	**Parameter settings**
Population size	200
Generation	60
Crossover probability	0.8
Mutation probability	0.01
Crossover method	Binary crossover
Mutation method	Binary mutation
Selection method	Tournament selection

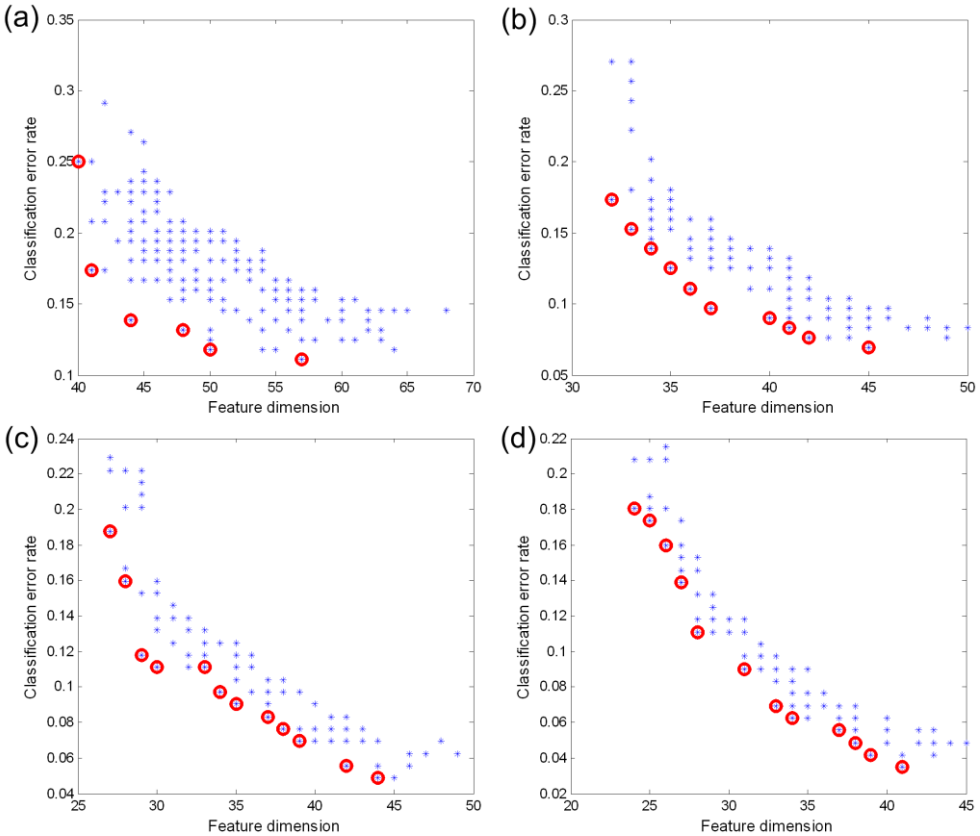

Fig. 8 Distributions of the solutions at different generations over the objective plane utilizing the NSGA-Ⅱ based on the original feature subset (114 features)

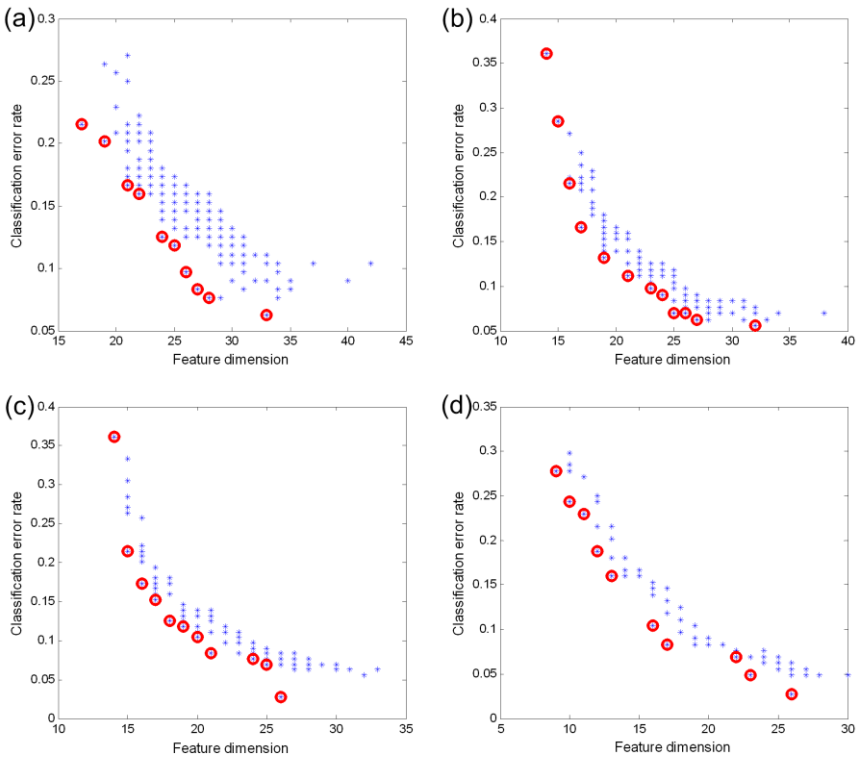

Fig. 9 Distributions of the solutions at different generations over the objective plane utilizing the NSGA-Ⅱ based on the ReliefF feature subset (59 features)

4.5 Comparison of the different feature selection methods

In this section, the classification performance of four different feature subsets, mean the original feature subset ($F_{Original}$), the feature subset obtained by filter method based on ReliefF ($F_{ReliefF}$), the feature subset obtained by wrapper method based on NSGA-Ⅱ (F_{NSGA}) and the feature subset obtained by the hybrid filter and wrapper method ($F_{ReliefF+NSGA}$), are evaluated and compared. The computation time, feature subset dimensions as well as the classification accuracy of different feature subsets based on three classifiers are shown in Table 4.

Table 4 Gear faults classification performances of four feature subsets $F_{Original}$, $F_{ReliefF}$, F_{NSGA} and $F_{ReliefF+NSGA}$ based on three classifiers

It can be found from Table 4 that the original feature subset $F_{Original}$, which consists of all the 107 features, obtained the worst classification performance by using all the three classifiers. It ascertains our assumption that there exist many irrelevant and redundant features, which will decrease the performances in the original feature subset. A feature selection procedure is indispensable before classification.

Table 4 Gear faults classification performances of four feature subsets $F_{Original}$, $F_{ReliefF}$, F_{NSGA} and $F_{ReliefF+NSGA}$ based on three classifiers

Classifiers	Feature Subsets	Dimension of the Feature subset	Computation time (s)	Classification accuracy (%)
KNNC	$F_{Original}$	114	--	72.92
	$F_{ReliefF}$	59	2.3	87.5
	F_{NSGA}	41	142.4	96.53
	$F_{ReliefF+NSGA}$	26	87.6	**97.23**
NBC	$F_{Original}$	114	--	68.75
	$F_{ReliefF}$	59	2.3	83.3
	F_{NSGA}	40	141.7	**94.89**
	$F_{ReliefF+NSGA}$	24	75.3	94.44
SVM	$F_{Original}$	114	--	75
	$F_{ReliefF}$	59	2.3	89.58
	F_{NSGA}	38	876.4	95.83
	$F_{ReliefF+NSGA}$	17	549.8	**98.61**

The performances of the feature subset $F_{ReliefF}$ showed to be better than the $F_{Original}$. Otherwise, the dimension of $F_{ReliefF}$ is nearly half of the original feature subset. However, the classification accuracies of $F_{ReliefF}$ are inferior to F_{NSGA} and $F_{ReliefF+NSGA}$ obviously.

The feature subset F_{NSGA} and $F_{ReliefF+NSGA}$, which are obtained by wrapper method and hybrid filter and wrapper method, have demonstrated similar classification accuracies in our case. The $F_{ReliefF+NSGA}$ achieved the highest classification rates by using KNNC and LS-SVM classifiers, while the F_{NSGA} gained the best performance by using NBC classifier. However, we can observe from Table 4 that the feature dimensions of $F_{ReliefF+NSGA}$ are lower than the F_{NSGA}. Moreover, the hybrid feature selection scheme required much less computation cost than the wrapper method. Therefore, it is very desirable to use hybrid filter and wrapper feature selection scheme to get a satisfactory refined feature subset for gear fault diagnosis.

5. CONCLUSION

This investigation has presented a hybrid filter and wrapper feature selection scheme for gear fault diagnosis based on ReliefF and NSGA-II algorithm. The filter method is implemented by

ReliefF and a candidate feature subset can be obtained by selecting the most outstanding features exceed a predefined threshold. Based on the candidate feature subset, wrapper technique combined with the multi-objective optimization evolutionary algorithm NSGA-II is adopted to get a more compact feature subset and higher classification accuracy.

We conducted experiments on a gearbox, in which four gear operating states were simulated, to evaluate the presented hybrid feature selection scheme. More than hundred parameters have been extracted as the original feature set by utilizing EMD, AR model, statistics, SVD and entropy techniques. Three popular classifiers were employed to evaluate the performances of the presented techniques. Moreover, some other feature selection schemes were also implemented and compared with the proposed approach.

Experimental results have revealed that the proposed hybrid filter and wrapper feature selection scheme achieves promising classification performance for gear fault diagnosis. The faults classification accuracies of three different classifiers using the selected features by the presented scheme were consistently higher than those using original feature subset and feature subsets obtained by other feature selection methods. The dimension of the feature subset obtained by the hybrid feature selection scheme is lower than the filter or wrapper methods. Furthermore, the computation cost of the hybrid feature selection is much less than the wrapper method.

It can be concluded that the proposed hybrid feature selection scheme demonstrates to be an attractive approach for gear fault diagnosis in both performance and efficiency. It can be easily extended to fault diagnosis for other mechanical machineries such as bearings and engines.

REFERENCES

[1] M. Amarnath and I.R.P. Krishna, Local fault detection in helical gears via vibration and acoustic signals using EMD based statistical parameter analysis, Measurement (2014), 154-164.

[2] J.L. Yin, W.Y. Wang, Z.H. Man and S.Y. Khoo, Statistical modeling of gear vibration signals and its application to detecting and diagnosing gear faults, Information Sciences (2014), 295-303.

[3] D.P. Jena and S.N. Panigrahi, Automatic gear and bearing fault localization using vibration and acoustic signals, Applied Acoustics (2015), 20-33.

[4] X.H. Liang, M.J. Zuo and M.R. Hoseini, Vibration signal modeling of a planetary gear set for tooth crack detection, Engineering Failure Analysis (2015), 185-200.

[5] B. Samanta, Gear fault detection using artificial neural networks and support vector machines with ge netic algorithms, Mechanical Systems and Signal Processing 3 (2004), 625-644.

[6] N. Saravanan, S. Cholairajan and K.I. Ramachandran, Vibration-based fault diagnosis of spur bevel gear box using fuzzy technique, Expert Systems With Applications 2 PART 2 (2009), 3119-3135.

[7] N. Saravanan, V.N.S.K. Siddabattuni and K.I. Ramachandran, Fault diagnosis of spur

bevel gear box using artificial neural network (ANN), and proximal support vector machine (PSVM), Applied Soft Computing 1 (2010), 344-360.

[8] G. Cheng, Y.L. Cheng, L.H. Shen, J.B. Qiu and S. Zhang, Gear fault identification based on Hilbert-Huang transform and SOM neural network, Measurement 3 (2013), 1137-1146.

[9] D.J. Bordoloi and R. Tiwari, Support vector machine based optimization of multi-fault classification of gears with evolutionary algorithms from time-frequency vibration data, Measurement (2014), 1-14.

[10] V. Havale and S. Narayanan, Diagnosis of manufacturing defects in a gear pair using wavelet analysis of vibration and acoustic signals and an ANN-based inference technique, Insight 8 (2014), 426-433.

[11] M. Khazaee, H. Ahmadi, M. Omid, A. Moosavian and M. Khazaee, Classifier fusion of vibration and acoustic signals for fault diagnosis and classification of planetary gears based on Dempster-Shafer evidence theory, Proceedings Of the Institution Of Mechanical Engineers Part E-Journal Of Process Mechanical Engineering 1 (2014), 21-32.

[12] D.L. Yang, Y.L. Liu, S.B. Li, X.J. Li and L.Y. Ma, Gear fault diagnosis based on support vector machine optimized by artificial bee colony algorithm, Mechanism And Machine Theory (2015), 219-229.

[13] G.G. Yen, Wavelet packet feature extraction for vibration monitoring, IEEE Transactions on Industrial Electronics 3 (2000), 650-667.

[14] Q. Hu, Z. He, Z. Zhang and Y. Zi, Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble, Mechanical Systems and Signal Processing 2 (2007), 688-705.

[15] Y. Lei, Z. He, Y. Zi and Q. Hu, Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs, Mechanical Systems and Signal Processing 5 (2007), 2280-2294.

[16] V. Sugumaran, V. Muralidharan and K.I. Ramachandran, Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing, Mechanical Systems and Signal Processing 2 (2007), 930-942.

[17] V. Sugumaran and K.I. Ramachandran, Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing, Mechanical Systems and Signal Processing 5 (2007), 2237-2247.

[18] L.B. Jack and A.K. Nandi, Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms, Mechanical Systems and Signal Processing 2-3 (2002), 373-390.

[19] Z. Chen, Y. He, F. Chu and J. Huang, Evolutionary strategy for classification problems and its application in fault diagnostics, Engineering Applications of Artificial Intelligence 1 (2003), 31-38.

[20] B. Samanta, K.R. Al-Balushi and S.A. Al-Araimi, Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection, Engineering Applications of Artificial Intelligence 7-8 (2003), 657-665.

[21] B. Samanta, K.R. Al-Balushi and S.A. Al-Araimi, Artificial neural networks and genetic algorithm for bearing fault detection, Soft Computing 3 (2006), 264-271.

[22] B. Samanta and C. Nataraj, Use of particle swarm optimization for machinery fault detection, Engineering Applications of Artificial Intelligence 2 (2009), 308-316.

[23] M. Danubianu, S.G. Pentiuc and D.M. Danubianu, Data Dimensionality Reduction for Data Mining: A Combined Filter-Wrapper Framework, International Journal of Computers Communications & Control 5 (2012), 824-831.

[24] J.M. Cadenas, M.C. Garrido and R. Martinez, Feature subset selection Filter-Wrapper based on low quality data, Expert Systems with Applications 16 (2013), 6241-6252.

[25] C. Peng, M.H. Wang, Y. Shen, H.Q. Feng and A. Li, Reconstruction and Analysis of Transcription Factor-miRNA Co-Regulatory Feed-Forward Loops in Human Cancers Using Filter-Wrapper Feature Selection, Plos One 10 (2013), 11.

[26] Z.Y. Hu, Y.K. Bao, T. Xiong and R. Chiong, Hybrid filter-wrapper feature selection for short-term load forecasting, Engineering Applications of Artificial Intelligence (2015), 17-27.

[27] I.K. Marko Robnik-Sikonja, Theoretical and Empirical Analysis of ReliefF and RReliefF, Machine Learning 1-2 (2003), 23-69.

[28] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation 2 (2002), 182-197.

[29] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Snin, Q. Zheng, N.C. Yen, C.C. Tung and H.H. Liu, The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 1971 (1998), 903-995.

[30] S.J. Loutridis, Damage detection in gear systems using empirical mode decomposition, Engineering Structures 12 (2004), 1833-1841.

[31] B. Liu, S. Riemenschneider and Y. Xu, Gearbox fault diagnosis using empirical mode decomposition and Hilbert spectrum, Mechanical Systems and Signal Processing 3 (2006), 718-734.

[32] J. Cheng, D. Yu, J. Tang and Y. Yang, Local rub-impact fault diagnosis of the rotor systems based on EMD, Mechanism and Machine Theory 4 (2009), 784-791.

[33] Y. Lei, M.J. Zuo and M. Hoseini, The use of ensemble empirical mode decomposition to improve bispectral analysis for fault detection in rotating machinery, Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 8 (2010), 1759-1769.

[34] Y. Lei, M.J. Zuo, Z. He and Y. Zi, A multidimensional hybrid intelligent method for gear fault diagnosis, Expert Systems With Applications 2 (2010), 1419-1430.

[35] C. Junsheng, Y. Dejie and Y. Yu, A fault diagnosis approach for roller bearings based on EMD method and AR model, Mechanical Systems and Signal Processing 2 (2006), 350-362.

[36] M. Sebban and R. Nock, A hybrid filter/wrapper approach of feature selection using

information theory, Pattern Recognition 4 (2002), 835-846.

[37] O. Uncu and I.B. Tu?rks?en, A novel feature selection approach: Combining feature wrappers and filters, Information Sciences 2 (2007), 449-466.

[38] Z. Zhu, Y.S. Ong and M. Dash, Wrapper-filter feature selection algorithm using a memetic framework, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 1 (2007), 70-76.

[39] M.C. Lee, Using support vector machine with a hybrid feature selection method to the stock trend prediction, Expert Systems with Applications 8 (2009), 10896-10904.

[40] N. Srinivas and K. Deb, Muiltiobjective Optimization Using Nondominated Sorting in Genetic Algorithms, Evolutionary Computation 3 (1994), 221~248.

[41] L.B. Jack and A.K. Nandi, Genetic algorithms for feature selection in machine condition monitoring with vibration signals, IEE Proceedings: Vision, Image and Signal Processing 3 (2000), 205-212.

[42] P.J. Grother, G.T. Candela and J.L. Blue, Fast implementations of nearest neighbor classifiers, Pattern Recognition 3 (1997), 459-465.

[43] R.R. Yager, An extension of the naive Bayesian classifier, Information Sciences 5 (2006), 577-588.

[44] J.A.K. Suykens and J. Vandewalle, Least squares support vector machine classifiers, Neural Processing Letters 3 (1999), 293-300.

[45] P.J. R.P.W. Duin, P. Paclik, E. Pekalska, D. de Ridder, D.M.J. Tax, S. Verzakov, PRTools4.1, A Matlab Toolbox for Pattern Recognition, Delft University of Technology (2007),

[46] J.D. Wu, M.R. Bai, F.C. Su and C.W. Huang, An expert system for the diagnosis of faults in rotating machinery using adaptive order-tracking algorithm, Expert Systems With Applications 3 PART 1 (2009), 5424-5431.