

Bearing fault classification based on conditional random field

Guofeng Wang*, Xiaoliang Feng and Chang Liu

Key Laboratory of Mechanism Theory and Equipment Design of Ministry of Education, Tianjin University, Tianjin, China

Received 15 June 2012

Revised 13 December 2012

Accepted 26 December 2012

Abstract. Condition monitoring of rolling element bearing is paramount for predicting the lifetime and performing effective maintenance of the mechanical equipment. To overcome the drawbacks of the hidden Markov model (HMM) and improve the diagnosis accuracy, conditional random field (CRF) model based classifier is proposed. In this model, the feature vectors sequences and the fault categories are linked by an undirected graphical model in which their relationship is represented by a global conditional probability distribution. In comparison with the HMM, the main advantage of the CRF model is that it can depict the temporal dynamic information between the observation sequences and state sequences without assuming the independence of the input feature vectors. Therefore, the interrelationship between the adjacent observation vectors can also be depicted and integrated into the model, which makes the classifier more robust and accurate than the HMM. To evaluate the effectiveness of the proposed method, four kinds of bearing vibration signals which correspond to normal, inner race pit, outer race pit and roller pit respectively are collected from the test rig. And the CRF and HMM models are built respectively to perform fault classification by taking the sub band energy features of wavelet packet decomposition (WPD) as the observation sequences. Moreover, K-fold cross validation method is adopted to improve the evaluation accuracy of the classifier. The analysis and comparison under different fold times show that the accuracy rate of classification using the CRF model is higher than the HMM. This method brings some new lights on the accurate classification of the bearing faults.

Keywords: Fault classification, conditional random field, rolling element bearing, hidden Markov model

1. Introduction

Rolling element bearings are considered as one of the critical mechanical components in industrial applications and their defects usually cause malfunction to some extent and even lead to failure of the machinery [12]. Therefore, the classification of rolling element bearing fault is an effective and necessary means to judge the bearing state and predict the lifetime so that some measures can be taken to avoid the occurrence of the catastrophic results. Up to now, many generative and discriminative methods such as multilayer feed forward [1], support vector machine [10,13], Gaussian mixture model [6] and hidden Markov model (HMM) [15] have been presented to build the relationship between the selected features and the corresponding bearing fault. Among these methods, HMM model is preferred in some cases because of its strong capability of capturing temporal dynamic characteristics of the non stationary signal. The main principle of HMM is that the feature data and the state data for certain kind of fault can be depicted by a sequential model in which the state recognition is achieved by calculating the maximum production

*Corresponding author: Guofeng Wang, Key Laboratory of Mechanism Theory and Equipment Design of Ministry of Education, Tianjin University, Tianjin 300072, China. Tel.: +86 022 2740 6951; Fax: +86 022 2740 6260; E-mail: gfwangmail@tju.edu.cn.

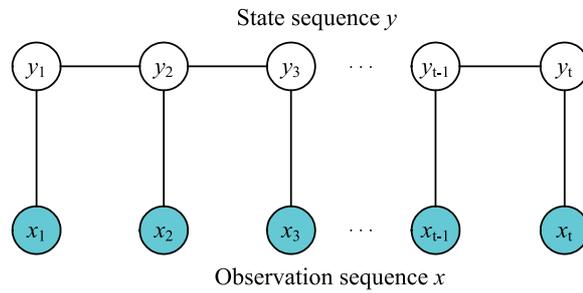


Fig. 1. Structure of conditional random field model.

probability of the feature sequence. Boutros and Liang [15] perform the classification of fault categories and its severity using the discrete HMM model. Purushothama et al. [17] build a wavelet based HMM model to depict the temporal information of the bearing vibration signal and perform the diagnosis of the multi faults. Lee et al. [9] propose a continuous HMM model to perform the diagnosis of mechanical faults. These applications show that the HMM based methods are effective to perform the classification of rolling element bearing fault. However, for a HMM model, the independence of the observation data needs to be assumed. Therefore it can't accommodate multiple overlapping features of the observation data or long range dependencies between the state indexes and the observation indexes at multiple time steps because the inference of HMM in this case becomes intractable [3,8]. While for the condition monitoring of the rolling element bearings, the sensory signal is usually split into arbitrary segments and these segments usually interrelate with each other because of their non stationary characteristics. In such case, the independence of the observation sequences extracted from these segments is hard to be satisfied, which will deteriorate the recognition accuracy of the HMM classifier correspondingly.

The conditional random field (CRF) model is presented in this paper to overcome the shortcomings of HMM. In comparison with the HMM, the CRF model is much more expressive because arbitrary dependencies on the observation sequence are permitted [8]. Therefore, it has stronger ability to incorporate a rich set of overlapping and non-independent features [3]. To analyze the effectiveness of the CRF model and compare with the HMM, both models are built to perform the classification of normal, inner race pit, outer race pit and roller pit. The vibration sensory signals are collected from the test rig and wavelet packet decomposition (WPD) is utilized to capture the temporal-frequency information and the energy features are calculated from the sub bands as the feature sequences. The analysis and comparison under different cross validation schemes show that the CRF model is more accurate than the HMM model. This method brings some new light on the online condition monitoring of the rolling element bearings.

2. Principle of linear chain CRF model

As shown in Fig. 1, the commonly used CRF model is an undirected graphical model in which the state sequence y is linked by edges in a linear chain.

Let $x = \{x_1, x_2, \dots, x_t\}$ and $y = \{y_1, y_2, \dots, y_t\}$ denote the observation sequence and the state sequence respectively, t is the length of observation data. For the observation data x , the probability of state sequence y can be depicted as [5]

$$P(y|x, \theta) = \frac{1}{Z(x)} \exp \left[\sum_k \lambda_k T_k(y_{i-1}, y_i, x, i) + \sum_m \mu_m S_m(y_i, x, i) \right] \tag{1}$$

Where, i is the index of element in the state sequence, k and m denote the index of the feature function. $Z(x)$ is the normalization factor which can be described as

$$Z(x) = \sum_y \exp \left[\sum_k \lambda_k T_k(y_{i-1}, y_i, x, i) + \sum_m \mu_m S_m(y_i, x, i) \right] \tag{2}$$

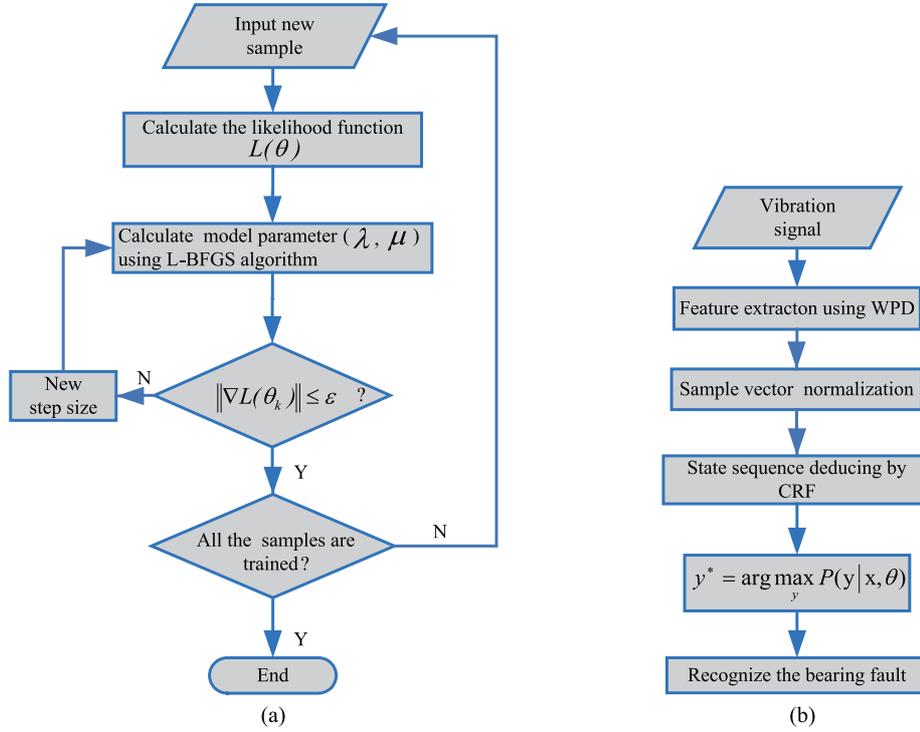


Fig. 2. Flowchart of CRF model parameters estimation and inference.

The state sequence y tends to satisfy the maximum global conditional probability [8] y^* , i.e.

$$y^* = \arg \max_y P(y|x) \tag{3}$$

The primary advantage of the CRF model is that it relaxes the independence assumptions of the HMM. Therefore, arbitrary dependencies on the observation sequence are allowed, which enhances the expressive ability of the classifier greatly.

For the application of linear-chain CRF model, the key problem is to solve the parameter vector θ . For the training data sets $D = \{ \langle x_t, y_t \rangle \}$, parameter estimation is performed by maximizing the log likelihood function which takes the form

$$L(\theta) = \sum_D \log[P(y_i|x_i, \theta)] \tag{4}$$

Where D denotes all the paired training data sequence, $x = \{x_1, x_2, \dots, x_t\}$ and $y = \{y_1, y_2, \dots, y_t\}$.

By combining Eq. (2) with Eq. (4), we get

$$L(\theta) = \sum_D \log \left\{ \frac{1}{Z(x)} \exp \left[\sum_k \lambda_k T_k(y_{t-1}, y_t, x_t, t) + \sum_m \mu_m S_m(y_t, x_t, t) \right] \right\} \tag{5}$$

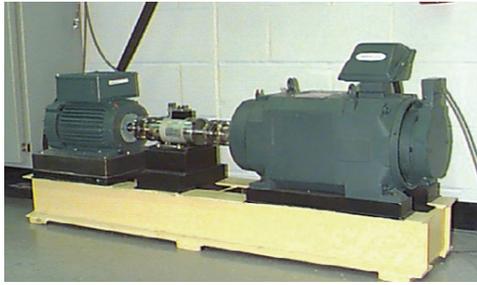
And the optimum estimation value $\hat{\theta}$ should satisfy

$$\hat{\theta} = \arg \max_{\theta} L(\theta) \tag{6}$$

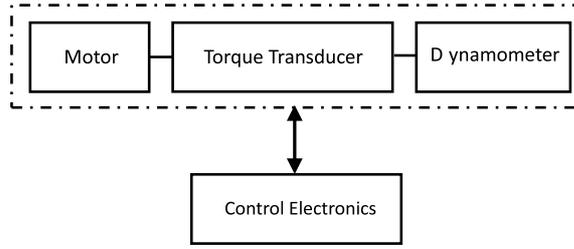
This function has the convexity property, which can ensure that the optimum model parameters can be calculated [8]. Given Eq. (6), Limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) [14] algorithm is adopted

Table 1
The parameters of the experimental bearings

Type	Inside diameter (mm)	Outside diameter (mm)	Thickness (mm)	Ball diameter (mm)	Pitch diameter (mm)
6205-2RS JEM SKF	25.00	52.00	15.00	7.94	39.04



(a) Picture of test rig



(b) Framework of test rig

Fig. 3. Scheme of test rig for rolling bearing.

here to find its maximum value and solve $\hat{\theta}$. The numerical optimization technique is adopted in this algorithm which makes it more efficient compared with the commonly used improved iterative scaling (IIS) method.

If the parameter $\theta = (\lambda_1, \lambda_2 \dots \lambda_k; \mu_1, \mu_2 \dots \mu_m)$ is available, the state recognition of the observation sequence can be performed using the forward backward algorithm. In this method, the forward probability $\alpha_i(y)$ is the probability of being state y given the observation sequence up to time i , which is described as [4]

$$\alpha_{i+1}(y) = \sum_{y'} \left\{ \alpha_i(y') \exp \left[\sum_k \lambda_k T_k(y', y, x, i+1) + \sum_m \mu_m S_m(y, x, i+1) \right] \right\} \quad (7)$$

y' represents all the possible state at time i . Similarly, $\beta_i(y)$ is the probability of being state y given the observation sequence after time i , which is represented as

$$\beta_i(y) = \sum_{y'} \left\{ \exp \left[\sum_k \lambda_k T_k(y, y', x, i+1) + \sum_m \mu_m S_m(y, x, i+1) \right] \beta_{i+1}(y') \right\} \quad (8)$$

Based on Eqs (7) and (8), the forward-backward and Viterbi algorithms [5] are utilized to deduce the state sequence correspondingly. The flowchart of the parameter estimation and model inference algorithm is given in Fig. 2.

3. Data preparation and feature extraction

3.1. Data preparation

The scheme of experimental rig is illustrated in Fig. 3. The test rig consists of a 2 hp motor (left), a torque transducer/encoder (center), a dynamometer (right), and control electronics (not shown). The deep groove ball bearings, whose detailed information is listed in Table 1, were utilized to support the motor shaft. An accelerometer was mounted on the drive end of the motor with magnetic base to acquire the vibration signals from the bearing under the rotating speed of 1797 RPM. Four kinds of faults, which are normal bearing, inner race pit, outer race pit and roller pit fault respectively, were simulated using the Electric Discharge Machining (EDM) method. Considering the vibration signal of the test rig is mainly focused on the low frequency band, the sampling frequency was selected as 12000 Hz. The length of each sample was selected as 2048, which cover more than five cycles of shaft rotating period. To ensure the generalization capability of the classifier, 100 samples were collected from every kind of bearing states, therefore the total number of the samples was 400 [7]. The typical waveform of four kinds bearing states is illustrated in Fig. 4. To further build the mapping relationship between the original waveform and the bearing faults, WPD based method is proposed in the next section to characterize the feature of the bearing signal and the sub bands energy is extracted as the observation sequence to build the classifier to perform the bearing fault classification.

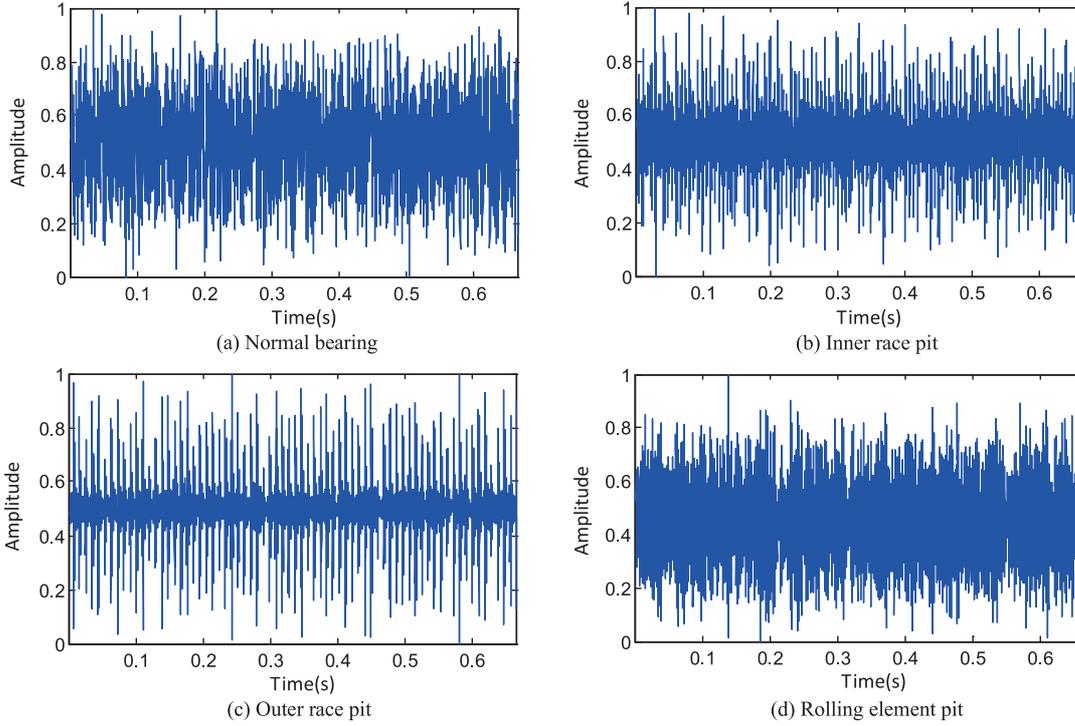


Fig. 4. Waveform of four kinds of bearing fault signals.

3.2. Feature extraction using WPD – wavelet packet decomposition

To depict the temporal and frequency characteristics of bearing vibration signal, WPD is adopted which is an extension of the wavelet decomposition (WD). The main characteristic of WPD is that it also decomposes the high frequency bands which are usually kept intact in the WD. Therefore, the time frequency information in the high frequency bands can also be depicted in details. The coefficients in the j th level for WPD can be written as [16]

$$\begin{cases} d_j^n(k) = \sum_m h_0(m - 2k)d_{j-1}^{n/2}(m) & (n \text{ is even}) \\ d_j^n(k) = \sum_m h_1(m - 2k)d_{j-1}^{(n-1)/2}(m) & (n \text{ is odd}) \end{cases} \quad (9)$$

Where, h_0 and h_1 are the filters related to the wavelet function, m is the subscript of $(j - 1)$ th level for WPD sub band and k is the subscript of j th level for WPD sub band. After the signal is decomposed by j levels, the energy in the n th sub band can be calculated as

$$E_j^n = \sum_k [d_j^n(k)]^2 \quad (10)$$

The frequency ranges of all sub bands at the j th level are

$$\left\{ \left[0, \frac{f_s}{2^{j+1}} \right]; \left[\frac{f_s}{2^{j+1}}, \frac{2f_s}{2^{j+1}} \right]; \dots; \left[\frac{(2^j - 1)f_s}{2^{j+1}}, \frac{f_s}{2} \right] \right\}$$

and f_s is the sampling frequency of the sensory signal. Here the level of the WPD is selected as 4, so the final dimension of the feature vectors is 8. The feature vector formed by all sub bands can be arranged as $\{E_4^1, E_4^2, E_4^3, \dots, E_4^8\}$. It should be noticed that these energy features need to be normalized firstly to get rid of the influence of the sample size. The spatial distribution of the wavelet packet energy features is demonstrated in

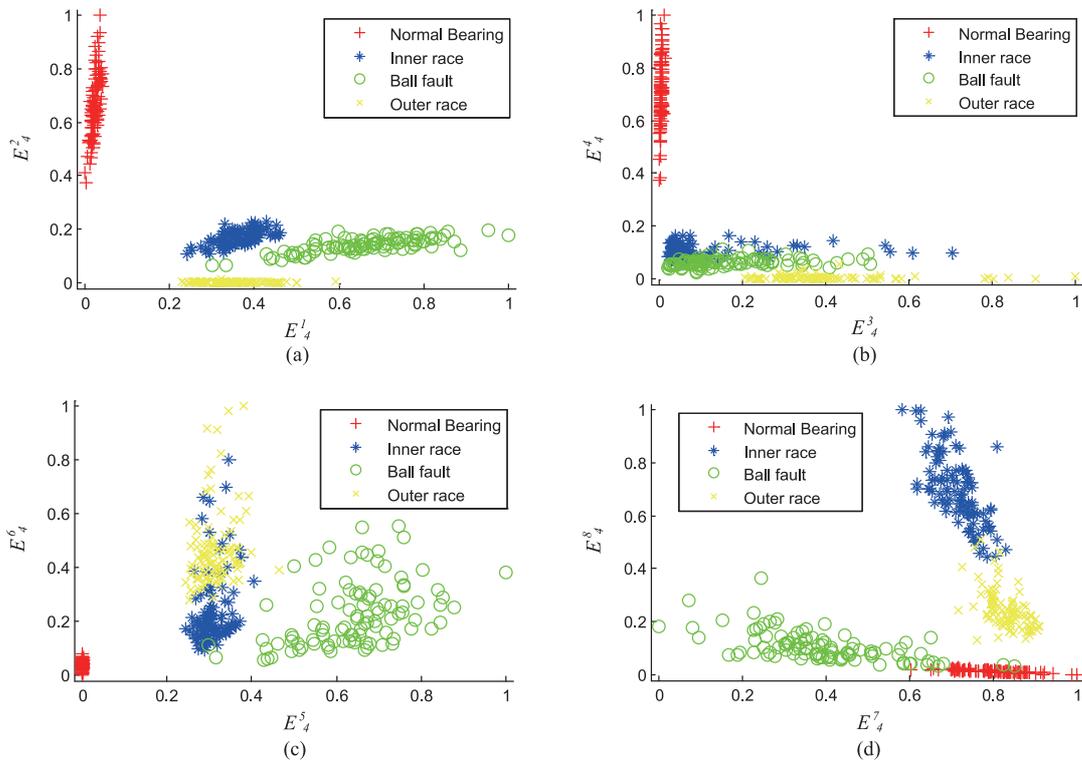


Fig. 5. The spatial distribution of the wavelet packet energy features (E_4^n denotes the n th sub band energy at level 4).

Fig. 5 from which it can be seen that the spatial distribution of the feature vectors is very disperse and its shape is also irregular. At the same time, some feature data under different fault states even mix with each other, which makes the boundary of each state not clear. In such case, the observation data demonstrate complicated temporal characteristics and it is not reasonable to judge the states of the bearing only by a single sample. Therefore, the sequence based pattern recognition model is adopted in the following section to judge the states of the rolling element bearings.

4. Bearing fault classification based on CRF

4.1. k -fold cross validation

Cross validation is one of the most commonly used performance assessment method [2]. In a k -fold cross validation, the training data are randomly split into k mutually exclusive subsets (the folds) of approximately equal size. The classifier is obtained by training $k - 1$ subsets and then tested using the remaining one subset. This process is repeated k times and each subset is utilized for testing only once. By calculating the averaging test errors over k trials, it gives an estimation of the expected generalization error. Here in this paper, two, five and ten folds are adopted respectively to evaluate the performance of the built classifier.

4.2. Classification based on CRF

The classification of the bearing fault using CRF model can be performed according to the following two steps. The first step is the training of the CRF model. As shown in Fig. 2(a), the model parameters in this stage are estimated based on the training sequences which are composed of wavelet packet energy features. The initial model parameters λ and μ are set to zero and the convergent accuracy limit ε is 0.0001. Based on the samples split according to the

Table 2
Average accuracy of different cross validation scheme

	Two fold	Five fold	Ten fold
HMM(%)	91.75	89.00	96.00
CRF(%)	100.00	98.75	99.75

Table 3
Average computational time of different cross validation scheme

	Two fold	Five fold	Ten fold
HMM(s)	25.44	31.12	25.82
CRF(s)	4.05	4.63	4.02

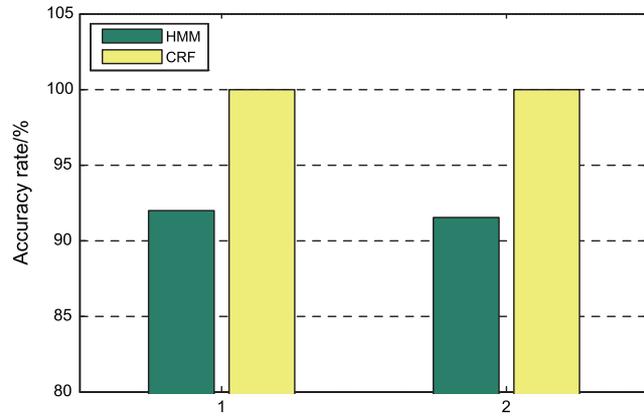


Fig. 6. Comparison of CRF with HMM under two-fold cross validation.

above cross validation schemes, the models can be built correspondingly using the L-BFGS algorithm. Different from HMM, only one model need to be built for the CRF classifier, which simplifies the training process of classifier greatly. The second step is bearing fault classification. The test samples are inputted into the CRF model and the predicted state sequence is calculated from the CRF model by forward backward inference. By comparing with the ideal state sequence, the accuracy rates are calculated under different cross validation schemes and demonstrated in the Figs 6–8. The average accuracy for different cross validation scheme is given in Table 2 from which we can see that the maximum accuracy can reach up to 100%.

4.3. Comparison with HMM

To make a comparison with CRF model, HMM is adopted simultaneously to perform the bearing fault classification with the same data. HMM [11] model is composed of Markov chain and stochastic process, which can be described as

$$\omega = \{N, M, \pi, A, B\} \quad (11)$$

The selection of N and M brings important effect on the accuracy rate of the classifier and there is no analytic method to calculate them accurately. Therefore, the trial and error method is commonly used to select them.

The probability of the test data coming from each HMM model is calculated by summing the probability of each hidden state [11].

$$P(O|\omega) = \sum_{i=1}^N P(O, q_t = i | \omega) \quad (12)$$

Where, O is the observation value of HMM model, N is the number of hidden states, q_t denotes the current state. The bearing fault category C is finally calculated by the following equation

$$C = \{z | P_z = \max[P(x|\omega_i)], i = 1, 2 \dots z\} \quad (13)$$

z is the serial number of the category corresponding to the maximum probability P_z . This equation means that the label corresponding to the maximum probability is chosen as the final fault category. In comparison with CRF model, the training process of HMM is complicated. The reason is that four models need to be trained respectively to recognize four kinds of fault states. Here, N is selected as 12 and M is equal to 8. After these models are built, the test samples are input to these models respectively and the accuracy rate is calculated correspondingly. The results are demonstrated in Figs 6–8 and the average accuracy for each kind of fault is listed in Table 2. It can be seen that the accuracy rate of HMM is lower than the CRF for all fold scheme used. To further compare the efficiency of two methods, the average computational time under different cross validation scheme are calculated using Intel core2 processor with 4 GB memory and the results are listed in Table 3. It can be seen that the CRF can get higher computational speed under different number of folds. Therefore, it proves that the CRF outperforms the HMM model for the classification of the rolling element bearing faults, which makes it more suitable for the industrial applications.

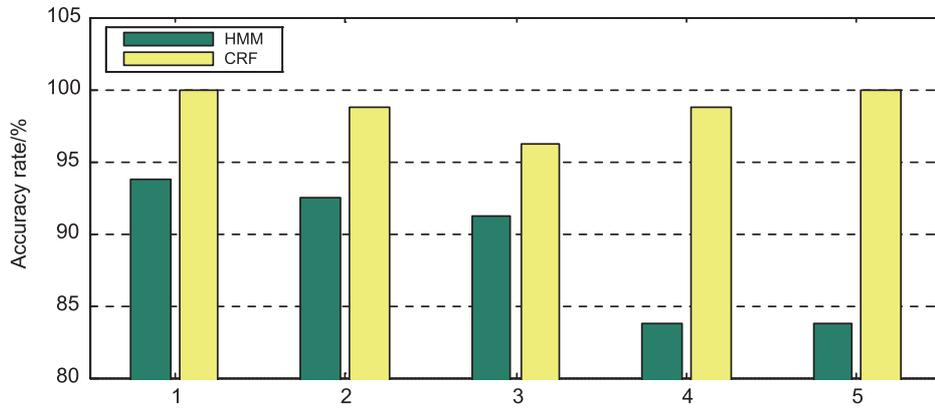


Fig. 7. Comparison of CRF with HMM under five-fold cross validation.

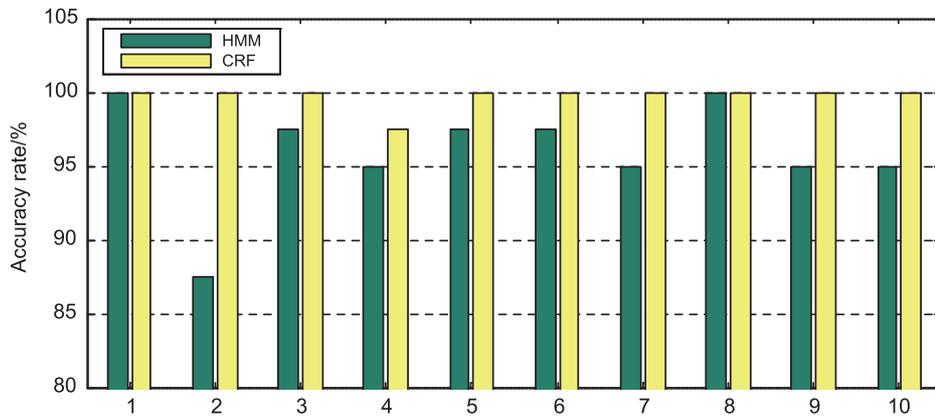


Fig. 8. Comparison of CRF with HMM under ten-fold cross validation.

5. Conclusions

In this paper, CRF is proposed to overcome the shortcomings of the HMM based classifier and perform the fault classification of the rolling element bearings. The main characteristic of the CRF model is its strong capability to depict the interrelationship between the observation sequences of different time steps by relaxing the independence hypothesis. Another advantage is that only one model needs to be trained for the multi fault classification, while the HMM needs to build one model for each kind of fault. Four kinds of bearing signal are utilized and the energy based features are extracted by wavelet packet decomposition to depict the relationship between the sensory signals and the bearing faults. The cross validation under different k folds proves that the CRF model has higher accuracy in comparison with the HMM. However, similar to HMM, the training of CRF needs relatively large sample data. Therefore, the training of CRF classifier under small dataset need to be further studied. In addition, because of the limit of the test rig, the current study is mainly focused on the single bearing faults under the stationary load regime. Therefore, the classification of the multiple bearing faults under varying load regime is also a promising research field.

Acknowledgment

The authors would like to thank Case Western Reserve University for their providing free download of the rolling element bearing fault data sets. This project is supported by National Natural Science Foundation of China (51175371).

Appendix: List of notation

A	State transfer probability matrix
B	Probability matrix of the observed values
C	Bearing fault category
$d_j^n(k)$	Coefficients in the j th level for WPD
$D = \{ \langle x_t, y_t \rangle \}$	Training dataset
E_j^n	Energy in the n th sub band of the j th level
f_s	Sampling frequency
h_0, h_1	Filters related to the wavelet function
$L(\theta)$	Log likelihood function
M	Possible number of the observed value in each state
N	State number of the Markov chain
O	Observation value of the HMM model
P_z	Maximum probability of the HMM model
$P(O \omega)$	Probability of each hidden state for the HMM model
$P(y x)$	Probability of state sequence for the CRF model
q_t	Hidden state of the HMM model
$S_m(y_i, x, i)$	State feature function of y_i and x
$T_k(y_{i-1}, y_i, x, i)$	Transition feature function of state y_i and y_{i-1} for the observation sequence x
$x = \{x_1, x_2, \dots, x_t\}$	Observation sequence
$y = \{y_1, y_2, \dots, y_t\}$	State sequence
y^*	Optimal sequence
z	Serial number of the bearing fault category
$Z(x)$	Normalization factor
$\alpha_i(y)$	Probability of being state y given the observation sequence up to time i
$\beta_i(y)$	Probability of being state y given the observation sequence after time i
$\theta = (\lambda_1, \lambda_2 \dots \lambda_k; \mu_1, \mu_2 \dots \mu_m)$	Parameters of the CRF model
$\hat{\theta}$	Optimum estimation value of θ
λ_k	Prior probability of the transition from the state y_i to y_j
μ_m	Prior probabilities of the state y_i
π	Initial probability distribution vector of HMM model
$\omega = \{N, M, \pi, A, B\}$	Parameters of the HMM model

References

- [1] 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* **16** (2003), 657–665.
- [2] C. Huang, Y. Lee, D. Lin and S. Huang, Model selection for support vector machines via uniform design, *Computational Statistics & Data Analysis* **52** (2007), 335–346.
- [3] C. Sminchisescu, A. Kanaujia and D. Metaxas, Conditional models for contextual human motion recognition, *Computer Vision and Image Understanding* **104** (2006), 210–220.
- [4] C. Sun, Y. Guan, X. Wang and L. Lin, Rich features based conditional random fields for biological named entities recognition, *Computers in Biology and Medicine* **37** (2007), 1327–1333.
- [5] C. Sutton and A. McCallum, An introduction to conditional random fields for relational learning, <http://www.cs.umass.edu/~mccallum/papers/crf-tutorial.pdf>, 2005.
- [6] G. Wang, Y. Li and Z. Luo, Fault classification of rolling bearing based on reconstructed phase space and Gaussian mixture model, *Journal of Sound and Vibration* **323** (2009), 1077–1089.
- [7] G. Wang, Z. Luo, X. Qin, Y. Leng and T. Wang, Fault identification and classification of rolling element bearing based on time-varying autoregressive spectrum, *Mechanical Systems and Signal Processing* **22** (2008), 934–947.
- [8] J. Lafferty, A. McCallum and F. Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, *Proceedings of the 18th International Conference on Machine Learning*, San Francisco, (2001), 282–289.
- [9] J. Lee, S. Kim, Y. Hwang and C. Song, Diagnosis of mechanical fault signals using continuous hidden Markov model, *Journal of Sound and Vibration* **276** (2004), 1065–1080.
- [10] 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* **25** (2012), 326–344.
- [11] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE* **77** (1989), 257–286.

- [12] N. Tandon and A. Choudhury, A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings, *Tribology International* **32** (1999), 469–480.
- [13] P. Konar and P. Chattopadhyay, Bearing fault detection of induction motor using wavelet and support vector machines (SVMs), *Applied Soft Computing* **11** (2011), 4203–4211.
- [14] R.H. Byrd, P. Lu and J. Nocedal, A limited memory algorithm for bound constrained optimization Ph.D. Dissertation, *Northwestern University Department of Electrical Engineering and Computer Science*, 1995.
- [15] T. Boutros and M. Liang, Detection and diagnosis of bearing and cutting tool faults using hidden Markov models, *Mechanical Systems and Signal Processing* **25** (2011), 2102–2124.
- [16] T. Wu, G. Yan, B. Yang and H. Sun, EEG feature extraction based on wavelet packet decomposition for brain computer interface, *Measurement* **41** (2008), 618–625.
- [17] V. Purushothama, S. Narayanana and S. Prasad, Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition, *NDT&E International* **38** (2005), 654–664.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

