

Fault diagnosis of rolling bearing based on PSO and continuous Gaussian mixture HMM

Liao Guangchun^{1, a}, Zhu Haiping^{1, b}, Liu Kangjun^{1, c} and Liao Jiawei^{1, d}

¹ Huazhong University of Science & Technology, Wuhan 430074, China;

^aglennliao@foxmail.com, ^bhaipzhu@hust.edu.cn, ^ckjliu@hust.edu.cn, ^dliaojiawei.leo@foxmail.com

Keywords: Particle Swarm Optimization; Hidden Markov Model; fault diagnosis; LPC

Abstract. As the hidden Markov model (HMM) has a strong ability of time sequence modeling, the continuous Gaussian mixture HMM is used to establish a model base of the rolling bearing fault. An adaptive particle swarm optimization (APSO) with extremum disturbed operator and dynamic change of inertia weights is introduced to the traditional training algorithm for solving the local extremum problem. The vibration signal is collected for extracting 12 order LPC coefficients as a feature vector through the dispose of adding window. In the given feature vector, the HMM is built for bearing fault condition monitoring and fault diagnosis. Then, different fault conditions experiment are carried out on the motor bearing test-bed. The experiment result shows that the method can use a small amount of samples for training HMM, and it is more effective and has higher classification accuracy in fault diagnosis compared with the traditional training algorithm.

1. Introduction

The rolling bearing is widely used in all kinds of rotating machinery; it is also one of the most vulnerable parts. As is known, the bearing failure will causes abnormal vibration and noise on running mechanical equipment, even leading to the equipment damage. Thus, an effective method of bearing fault diagnosis, identifying the degraded conditions of bearing correctly, plays a positive role in avoiding mechanical accident. Hidden Markov model (HMM) has a strong sequential model building ability of signal processing, which is very suitable for analyzing the nonstationary signal with poor reproducibility. HMM has successfully solved the problem of speech recognition and become one of main methods in the field of speech recognition. HMM is a dual stochastic process that its transfer between different states is a Markov process and the relationship between the state and its corresponding observed values is another stochastic process. The state of model cannot be observed directly, which is a kind of hidden state. It should be estimated through the observation symbols. Fault diagnosis of mechanical equipment is extremely similar to HMM. The status of equipment is not a direct observation which is judged by the runtime parameters of mechanical equipment. Considering the similarity of the mechanical vibration signal and the speech signal, scholars at home and abroad have applied HMM to the field of fault diagnosis. Bunks [1] pointed out that the method based on HMM is available for mechanical equipment state prediction. Hasan, etc. [2] trained the HMM via the historical data of bearing failure, and researched the fault diagnosis and prediction technology of bearing. Singular spectrum analysis and continuous hidden Markov model are used for bearing fault diagnosis and performance assessment in [3].

Traditional training algorithm of HMM has the disadvantage of slow convergence and falling into local minimum, that has significantly affected its performance in bearing fault diagnosis. This paper introduces an adaptive swarm optimization (APSO) algorithm, and applies it to the training of HMM. A new method of HMM training based on APSO algorithm is proposed, which effectively improves the capability of HMM in fault identification.

2. Proposed Method

2.1. Hidden Markov Model

In general, a HMM is usually defined by the following parameters:

- The number of states in HMM, N .
- The number of observation corresponding to each state, M .
- The initial probability distribution vector, $\pi = \{\pi_i\}$, where $\pi_i = P(q_1 = \theta_i), 1 \leq i \leq N$.
- The state transferring probability matrix, $A = (a_{ij})_{N \times N}$ where $a_{ij} = P(q_{t+1} = \theta_j | q_t = \theta_i), 1 \leq i, j \leq N$ which means the probability of being in state j at time $t+1$ when the state at time t is i .
- The observation probability matrix, $B = (b_{jk})_{N \times M}$ where $b_{jk} = P(o_t = v_k | q_t = \theta_j), 1 \leq j \leq N, 1 \leq k \leq M$ which means the probability of the k th observation being observed when the state at time t is j . If the observation is modeled as continuous, each state has a different set of probability density function.

The shorthand notation $\lambda = (\pi, A, B)$ can be used to express a HMM.

In the rolling bearing fault diagnosis, most of scholars' research use the discrete HMM for classifying faults of the rolling bearing [4][5]. However, the continuous HMM directly regards the observed values as observation sequence, and a discrete HMM need to firstly quantify the observed values. Therefore, the use of continuous HMM has the advantages of distortion and classification. This paper uses the continuous Gaussian mixture HMM for studying the fault diagnosis. The degradation of bearing is an irreversible process, a state can't return to the previous state. So the left-to-right Markov chain with five states (normal running and 4 different fault conditions) is chosen. A five-state HMM is shown in Fig.1. The initial probability distribution vector is set to $[1, 0, 0, 0, 0]$. The observation probability of each state can be represented by mixture Gauss density functions [6]. Observed values probability density function can be expressed as

$$b_j(X) = \sum_{k=1}^K c_{jk} N(X, \mu_{jk}, \Sigma_{jk}), 1 \leq j \leq N \quad (1)$$

where $N(X, \mu_{jk}, \Sigma_{jk})$ is a multidimensional Gaussian probability density function, X is the observation sequence, μ_{jk} is the average vector, Σ_{jk} is the covariance matrix, K is the number of composition in mixture Gauss density functions, c_{jk} is the combination coefficient and $\sum_{k=1}^K c_{jk} = 1$.

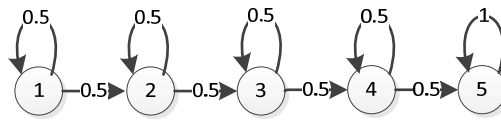


Fig. 1. Five-state left-to-right HMM

For purpose of strengthening the stability of HMM in fault diagnosis, multiple observation sequences are used for modeling and training the HMM [7]. A set of L observation is $O^{(l)} (l = 1, 2, \dots, L), O^{(l)} = \{o_1^{(l)}, o_2^{(l)}, \dots, o_T^{(l)}\}$. Assuming each observation sequence is independent, the probability of observation sequences set can be expressed as

$$P(O | \lambda) = \prod_{l=1}^L P(O^{(l)} | \lambda) \quad (2)$$

Revaluation equations for training HMM based on Baum-Welch algorithm are amended as

$$\bar{\pi}_i = \sum_{l=1}^L \alpha_1^{(l)}(i) \beta_1^{(l)}(i) / P(O^{(l)} | \lambda), 1 \leq i \leq N \quad (3)$$

$$\bar{a}_{ij} = \frac{\sum_{l=1}^L \sum_{t=1}^{T_l-1} \alpha_t^{(l)}(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}^{(l)}(j) / P(O^{(l)} | \lambda)}{\sum_{l=1}^L \sum_{t=1}^{T_l-1} \alpha_t^{(l)}(i) \beta_t^{(l)}(i) / P(O^{(l)} | \lambda)}, 1 \leq i, j \leq N \quad (4)$$

$$\bar{b}_{jk} = \frac{\sum_{l=1}^L \sum_{t=1 \cup \{o_t=v_k\}}^{T_l-1} \alpha_t^{(l)}(j) \beta_t^{(l)}(j) / P(O^{(l)} | \lambda)}{\sum_{l=1}^L \sum_{t=1}^{T_l-1} \alpha_t^{(l)}(j) \beta_t^{(l)}(j) / P(O^{(l)} | \lambda)}, 1 \leq j \leq N, 1 \leq k \leq M \quad (5)$$

In order to detect the presence of bearing failure, it is sufficient to train a single HMM corresponding to bearing in normal operating condition. The HMM is trained with observation sequence extracted from the collected vibration signal. Then, the probability of observation sequence for the normal condition can be calculated. Comparing with the predetermined threshold, if the probability is greater than the threshold means that there is no fault present in the bearing. Otherwise, the bearing is operating in fault condition. After detection of the bearing fault, fault diagnosis program is started. To achieve the goal of identifying correctly in which degradation state the bearing is. It is necessary to train HMMs to present its corresponding bearing faults. Similarly, the collected vibration signals are extracted to obtain observation sequence. The probability of observation sequence is calculated given all the HMMs in the previously established model base. The HMM, for which the probability is the largest, determines the operating condition of the bearing. The process of bearing fault diagnosis is shown in Fig.2.

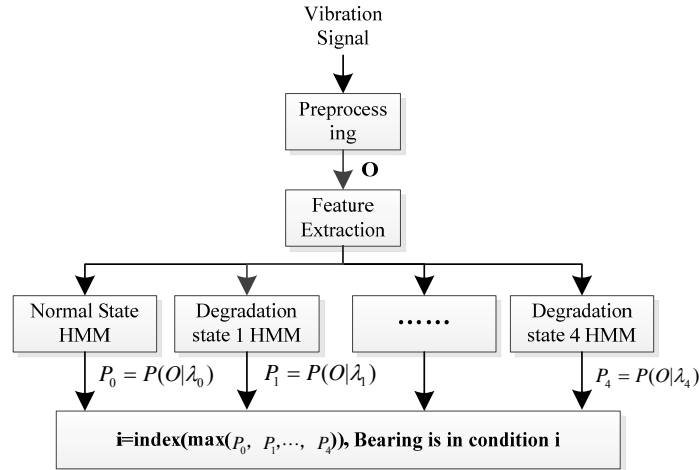


Fig. 2. Bearing fault diagnosis process diagram based on the HMM

2.2. Adaptive particle swarm optimization algorithm with extremal disturbance

Baum-Welch algorithm is relatively successfully solved the problem of training HMM parameters. But the algorithm belongs to the hill-climbing algorithm of local optimal, which depends on the initial value of parameters strongly. Therefore, the inaccurate estimation of initial value often leads to premature convergence and achieve local optimal solution. Particle swarm optimization (PSO) is a swarm intelligence technique based on group collaboration, which searches through the total solution space by the evolution of the particles. And it is more advantageous than the hill-climbing algorithm to find global optimal solution.

The basic PSO algorithm is proposed by Kennedy et al. [8] through the study of birds prey activities. Suppose that a particle population have N particles, including that the position of the i th ($i=1,2,\dots,N$) particle is expressed as $X_i = (x_i^1, x_i^2, \dots, x_i^D)$ and the speed of the i th ($i=1,2,\dots,N$) particle is expressed as $V_i = (v_i^1, v_i^2, \dots, v_i^D)$. Then the fitness $f(x)$ should be introduced into the optimization function. The position and speed of the particle are updated according to the following evolution equation as

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (6)$$

$$v_i(t+1) = wv_i(t) + c_1r_1(p_i(t) - x_i(t)) + c_2r_2(p_g(t) - x_i(t)) \quad (7)$$

where c_1 and c_2 are acceleration coefficients, w is inertia weight, r_1 and r_2 are random numbers between $[0, 1]$. $p_i(t)(i=1,2,\dots,N)$ denotes the optimal position of the i th particle in current search, $p_g(t)(i=1,2,\dots,N)$ denotes the optimal position of the particle population.

According to the research questions, iterative algorithm termination conditions may be selected to be greater than the maximum number of iterations or the optimal position of particle population in the current search can satisfy the demands of setting fitness.

1) Extremum disturbed operator

To solve the premature convergence problem of the PSO, disturbance operator is used to adjust the optimal positions of the particle and the particle population in this paper. With the adjustment, the new algorithm can escape from the local minimum of early convergence. Meanwhile, the accuracy of the solution is constantly improved and the disturbance operator becomes smaller along with the iteration. Disturbance operator are defined as follows:

$$\begin{cases} r_3 = (1-t/t_{\max})(1+(-2)*rand)+1 \\ r_4 = (1-t/t_{\max})(1+(-2)*rand)+1 \end{cases} \quad (8)$$

$$p_i = r_3 p_i, p_g = r_4 p_g \quad (9)$$

where t is the current number of iterations, t_{\max} is the largest number of iterations, then $(1-t/t_{\max})$ will gradually decreases to zero with the increasing the number of iteration. $rand$ is a random number between the $[0, 1]$, then $(1+(-2)*rand)$ generates the random number between $[-1, 1]$, which implements the disturbance on both sides of the extreme value point. So $0 < r_3, r_4 < 2$, and finally tend to 1. Disturbance operator is introduced into the equation (7) to generate a new update equation:

$$v_i(t+1) = wv_i(t) + c_1 r_1 (r_3 p_i(t) - x_i(t)) + c_2 r_2 (r_4 p_g(t) - x_i(t)) \quad (10)$$

2) Adaptive adjustment of inertia weight

Shi and Eberhard's research [9] shows that the different inertia weight value w has an obvious influence on the performance of the algorithm. On the one hand, the larger w can effectively avoid falling into local extremum and strengthen the global search ability of the algorithm; on the other hand, the smaller w is able to accelerate the convergence rate and is conducive to local search. This paper adopts an adaptive parameter setting method, w is defined as a linear function decreases along with the increase of the number of iterations. The linear function can be expressed as

$$w = w_{\max} - \frac{t * (w_{\max} - w_{\min})}{t_{\max}} \quad (11)$$

where $w_{\max}=0.9$ is a maximum of inertia weight, $w_{\min}=0.4$ is a minimum of inertia weight.

When equation (10) and (11) are introduced into the PSO algorithm, an new adaptive particle swarm optimization (APSO) algorithm is proposed in this paper.

3. The HMM Training Method Based on APSO Algorithm

To absorb the advantages of APSO algorithm in global optimization, the HMM training method based on APSO algorithm (APSO-HMM) incorporate APSO into Baum Welch algorithm in the process of the HMM training. A particle of the particle population represents a HMM in the new method for parameter optimization of HMM. First of all, APSO algorithm is used for global optimization under the solution space. Then, the Baum-Welch algorithm is used to do k times local search for the best HMM in every generation of the evolution by APSO algorithm. This method makes the traditional training methods to reduce the dependence on the initial value of parameters, and is conducive to jump out of local extremum. For continuous Gaussian mixture density HMM, this paper uses the new algorithm to optimize the parameters of HMM including the state transition probability matrix A and the observation probability matrix B . Among them, the observation probability matrix B consists of the combination coefficient c_{jk} , the average vector μ_{jk} and the covariance matrix Σ_{jk} . Each of these three elements above needs to be optimized so as to complete the parameter optimization of the observation probability matrix B . The detailed process of the training algorithm is shown in Fig.3.

```

do
  for i = 1 to N //N is the particle number of the population
     $x_i = \text{Baum-Welch}(x_i, k)$  //Each particle do k times local optimization
    if  $f(x_i) > f(p_i)$ 
       $p_i = x_i$ 
    end
    if  $f(x_i) > f(p_g)$ 
       $p_g = x_i$ 
    end
  for j = 1 to P //P is the number of the HMM parameters in the optimization
     $w = w_{\max} - \frac{t * (w_{\max} - w_{\min})}{t_{\max}}$ 
     $\{r_3 = (1 - t / t_{\max}) \times (1 + (-2) * \text{rand}) + 1$ 
     $\{r_4 = (1 - t / t_{\max}) \times (1 + (-2) * \text{rand}) + 1$ 
     $v_i(t+1) = wv_i(t) + c_1 r_1 (p_i(t) - x_i(t)) + c_2 r_2 (p_g(t) - x_i(t))$ 
     $x_i(t+1) = x_i(t) + v_i(t+1)$  // New particles needs to be normalized
  end
end
while(Did not meet the terminating conditions)

```

Fig. 3. The process of APSO-HMM training algorithm

4. The Experiment and Result Analysis

To verify the validity of the APSO-HMM algorithm, vibration data collected in the motor bearing experiment of Case Western Reserve University Bearing Data Center was used for experimental analysis [10]. The test platform consisted of a 2 horsepower electric motor, a torque sensor and a dynamometer. Vibration signal was collected using acceleration sensors, which were attached to the drive end of the motor housing at the 12 o'clock position. Vibration signals were recorded by a 16 channel DAT recorder, and post processed in MATLAB software.

In the experiment, the sampling frequency of the signal was 12 kHz. Vibration data was collected for normal running state and 4 different fault conditions. Faults were introduced into drive-end bearings using electro-discharge machining with fault diameters of 7 mils, 14 mils, 21 mils, and 28 mils(4 different fault conditions: degradation state 1, degradation state 2, ..., degradation state 4). The length of time for each vibration signal is 10 seconds. Before feature extraction the vibration signal was divided into T windows of equal length. By using the autocorrelation method, each window was encoded as a feature vector, which consisted of a set of 12 order LPC coefficients for that window. All the feature vectors for T windows were joined together to form an observation sequence, which will be used for training HMMs. The Feature process extraction is shown in Fig.4.

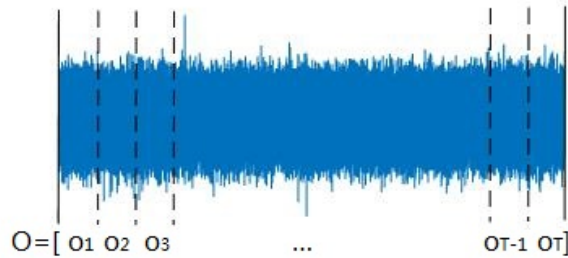


Fig. 4. Feature extraction of vibration signal

In the process of model training, the vibration signals for normal condition and 4 different fault conditions are divided into 20 equal segments, respectively. Each section is 0.5s long. Among them, the former 2 sections is used for training the HMMs, namely it takes $L=2$ observation sequences for modeling according to the equation (2). And the remaining 18 sections are used to verify the performance of the model. For feature extraction, the window size is set to 0.05s to extract the LPC coefficients as a feature vector. The initial state of transition probability matrix and continuous mixed Gaussian probability density function is random initialization. The parameters of APSO-HMM method for training HMM is set as follow: maximum number of iterations of the algorithm is 10, the particle number of the population is 10, the state number of HMM is 5, and Baum-Welch algorithm implements 4 times local optimization after each update of particle iteration. The parameters of

traditional Baum-Welch method for training HMM is set as follow: the termination condition is that the convergence error of the likelihood probability between two iterations is less than $e = 10^{-4}$, and the state number of HMM is 5. Each method test 5 times in the HMM training.

The average likelihood probability of the HMM training for APSO-HMM method and traditional Baum-Welch method in the tests is shown in Table 1. The greater the likelihood probability value of the HMM training indicates that the training of HMM is better. And its recognition ability is more prominent. The table shows that APSO-HMM algorithm is more excellent than traditional Baum-Welch algorithm in the model training.

Table 1 The average likelihood probability of HMM training

Training data	Average likelihood probability	
	APSO-HMM algorithm	Baum-Welch algorithm
normal	529.4603	521.2667
degradation 1	550.9924	540.1592
degradation 2	534.2993	505.0045
degradation 3	626.6712	572.0256
degradation 4	565.0951	509.8068

In the fault diagnosis process, a HMM model base, including the HMMs of five different operation state, is established through the use of APSO-HMM algorithm. Then, test sample data is input to the model base for fault diagnosis test. Table 2 shows the part of the logarithmic likelihood probability of test samples, which are calculated by the trained models. And diagnosis result shows that the 10 samples of test data get the correct classification.

Table 2 The logarithmic likelihood probability and diagnosis results

Test data	Logarithmic likelihood probability					Diagnosis result
	normal	degradation 1	degradation 2	degradation 3	degradation 4	
normal	207.9	-2544.6	-1883.6	-Inf	-3155.8	normal
normal	207.2	-2387.9	-1927.5	-7595.6	-2805.9	normal
degradation 1	-1840.9	243.9	164.0	-4005.8	-397.2	degradation 1
degradation 1	-1960.8	239.5	152.8	-Inf	-546.6	degradation 1
degradation 2	-1854.0	-72.4	182.4	-3784.2	-611.8	degradation 2
degradation 2	-1709.7	73.3	208.7	-4438.1	-518.8	degradation 2
degradation 3	-1943.2	-361.5	84.4	336.2	-635.9	degradation 3
degradation 3	-2067.0	-448.9	-5.8	224.4	-563.3	degradation 3
degradation 4	-3136.0	-1187.7	-1456.2	-5352.2	118.5	degradation 4
degradation 4	-2969.7	-1032.0	-1347.2	-5775.2	141.2	degradation 4

The remaining 18 sections of data is input to the HMM model base trained with APSO-HMM algorithm and the HMM model base trained with Baum-Welch algorithm for fault diagnosis test, respectively. The diagnosis result of the two model bases is shown in Table 3 and Table 4 below. The correct diagnosis for APSO-HMM algorithm is 72 times, and the accuracy of the classification is 100%. By contrast, the correct diagnosis for Baum-Welch algorithm is 70 times, the wrong diagnosis is 2 times, and the accuracy of the classification is 97.2%. Thus, it is confirmed that the HMM training method based on APSO algorithm has higher ability of fault identification.

Table 3 Bearing fault diagnosis of APSO-HMM

Test data	Test times	Normal	Degradation 1	Degradation 2	Degradation 3	Degradation 4
normal	18	18	0	0	0	0
degradation 1	18	0	18	0	0	0
degradation 2	18	0	0	18	0	0
degradation 3	18	0	0	0	18	0
degradation 4	18	0	0	0	0	18

Table 4 Bearing fault diagnosis of Baum-Welch

Test data	Test times	Normal	Degradation 1	Degradation 2	Degradation 3	Degradation 4
normal	18	18	0	0	0	0
degradation 1	18	0	16	2	0	0
degradation 2	18	0	0	18	0	0
degradation 3	18	0	0	0	18	0
degradation 4	18	0	0	0	0	18

5. Conclusion

In order to overcome the shortcomings of traditional algorithm in training HMM, this paper proposes a training method of HMM based on APSO algorithm (APSO-HMM). The experiment comparison of two training algorithms shows that the new training method has more excellent model training effect, and is better to complete the establishment of the HMM model base. Result Analysis shows that the HMM trained with APSO-HMM method has a higher recognition rate than the HMM trained with Baum-Welch algorithm. The new training algorithm can not only avoids falling into local minima, but also improves the ability of fault diagnosis.

Acknowledgment

The authors greatly acknowledge the financial supports from the National Science and Technology Major Project of China under the Grant No. 2014ZX04002011 and 2014ZX04014101.

References

- [1] Bunks, Carey, M. Dan, and T. Al-Ani. "Condition-based maintenance of machines using Hidden Markov Models." *Mechanical Systems & Signal Processing* 14.4(2000):597-612.
- [2] Ocak, Hasan, and K. A. Loparo. "HMM-Based Fault Detection and Diagnosis Scheme for Rolling Element Bearings." *Journal of Vibration & Acoustics* 127.4(2005):299-306.
- [3] Liu, T., J. Chen, and G. Dong. "Singular spectrum analysis and continuous hidden Markov model for rolling element bearing fault diagnosis." *Journal of Vibration & Control* 21(2013).
- [4] Feng Changjian, Kang Jing, et al. "Application in Fault Diagnosis of Rotary Machine Based on Theory Of DHMM Dynamic Pattern Recognition." *Journal of Dalian Nationalities University* 8.3(2006):12-15.
- [5] Boutros, Tony, and M. Liang. "Detection and diagnosis of bearing and cutting tool faults using hidden Markov models." *Mechanical Systems & Signal Processing* 25.6(2011):2102–2124.
- [6] Lee, Jong Min, et al. "Diagnosis of mechanical fault signals using continuous hidden Markov model." *Journal of Sound & Vibration* 276.3(2004):1065–1080.
- [7] Davis, R. I. A., B. C. Lovell, and T. Caelli. "Improved estimation of hidden Markov model parameters from multiple observation sequences." *International Conference on Pattern Recognition The Institute of Electrical and Electronics Engineers*, 2002:168-171.
- [8] Kennedy, James. "Particle swarm optimization." *Neural Networks, 1995. Proceedings., IEEE International Conference on IEEE*, 1995:1942-1948 vol.4.
- [9] Eberhart, Russ C., and Yuhui Shi. "Comparing inertia weights and constriction factors in particle swarm optimization." *Evolutionary Computation, 2000. Proceedings of the 2000 Congress on. Vol. 1. IEEE*, 2000.
- [10] The Case Western Reserve University Bearing Data Center. Bearing data center seeded fault test data[EB/OL].[2015-4-8].<http://csegroups.case.edu/bearingdatacenter/pages/download-data-file/>