Indian Journal of Science and Technology, Vol 9(10), DOI: 10.17485/ijst/2016/v9i10/85303, March 2016 ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Vibration based Health Assessment of Bearings using Random Forest Classifier

R. Satishkumar* **and V. Sugumaran**

School of Mechanical and Building Sciences, VIT University, Chennai Campus, Chennai - 600127, India; cr_sathi@yahoo.co.in, sugumaran.v@vit.ac.in

Abstract

Objective: This paper proposes a predictive model to assess the health condition of bearing using classification technique. **Method:** In the present study, vibration signals were acquired on a daily basis until the bearing is damaged. Initially, feature selection was done with decision tree and predictive model was built using selected features. Now, Random forest classifier was used to build the model to assess the remaining lifetime of the bearing. Distinct data were used to validate the performance of the classifier. **Findings:** The classification accuracy of the built model was found to be 95.64%. **Applications:** The proposed model was tested with the data acquired from a bearing experimental set-up wherein run-to-failure test were conducted on bearings at rated load conditions.

Keywords: Bearings, Life Time Assessment, Random Forest Classifier

1. Introduction

Any unplanned failure and breakdowns in the bearings may lead to loss of time, energy resources and manpower. Hence, assessing the remaining useful lifetime of bearing helps in saving energy resources and preventive measures can be taken. In the experimental setup, a brand new bearing was taken and vibration signals were acquired on a continuous basis till it fails naturally.

Most of the previous works were based on simulated faults which include inner race faults, outer race faults, cage faults, etc. and some works were based on natural faults. In order to save time, many research scholars used an accelerated load test in predicting the lifetime of bearing. However, the experiment which is conducted at rated load and speed conditions was different from accelerated load test as the phenomenon of degradation was different. In the present study, the vibration signals were acquired at rated load and speed conditions.

This research work was based on classification approach and predictive model was built to assess the remaining useful time of bearings. Vibration signals acquired from bearing were collected and statistical features were extracted. Decision tree was used to select the most contributing features. In the present study, out of 12 statistical features, 5 best features were selected for building the model. Later, with the selected features the predictive model was built using random forest classifier. The classification accuracy of the model was presented in results and discussion.

In the previous research works, the standards for predicting the lifetime of bearings at rated load and speed conditions were dealt[1,2]. Real time data, such as operational work (breakdowns, scrap, etc.), maintenance data were used in building the model. To assess the remaining lifetime of bearings, experience-based prognostic method was used. This method uses the basic reliability functions and it is simple to use. However, the results from this method were precise compared to other methods.

Acoustic signals or vibration signals will be acquired from the bearings to predict the lifetime of bearings. A comparative study of acoustic monitoring and conventional vibration was presented to identify defects in industrial multistage gear box, operating under faulty and healthy conditions[3]. The statistical parameter estimation

*Author for correspondence

method was presented to detect defects in rolling element bearing using sound pressure and vibration signals. This study concluded that kurtosis and crest factor from both sound and vibration signals provide better diagnostic information than beat function parameters[4]

Summarized dot pattern method was presented in the diagnosing bearing of the fan using sound signals. A diagrammatic representation present in the model helps in identifying the faults present in the bearing[5]. In all the above works, acoustic signals were acquired from the bearing to diagnose the defect present in it. However, in the present study vibration signals were used to assess the remaining useful time of the bearings.

Bearing maintenance is very essential in rotating machines as it directly affects the performance. They can be grouped into prognosis and diagnosis. The present state (safe stateor damaged state) of the bearing can be identified through diagnosis. Three different methods viz neural network[6,7], vibration analysis and statistical study[8–12] were deployed.

Stress based fatigue prognostic method was deployed to predict the bearing lifetime precisely. As the degradation phenomenons are not consistent over the period of time, it is highly challenging to build such a model. Hence, in real time applications stress based fatigue prognostic method is complex to build[13]

The prognosis is mainly focused on predicting the remaining lifetime of bearings. In two different ways prognosis was carried out, direct approach for fracture mechanics[14,15] and statistical approach of vibration signals[16]

In prediction of remaining useful time of bearings, data driven methods were used. They capture signals with accelerometers, sensors, etc. which are converted into data for the study of degradation of the system. Statistical method, spall propagation model and Artificial Neural Networks (ANN) were used to build the model. The built model quickly responds to the change in environment and accommodates to variable conditions. However, prognostic results from this method were precise and may not be suitable in all applications[17,18]

Bearing performance directly affects the machineries and hence, maintenance is essential to save human effort and energy resources. Fault diagnosis of bearings and prognosis seeks its importance in the last three decades due to its inevitable role in industrial applications. It is highly challenging to collect huge data on regular interval due to want of resources, time and money. The present work is based on classification approach to build the predictive model to assess the remaining useful time of bearing. Feature selection is done through decision tree and random forest classifier was used to build the proposed model. The outcome of the model was presented in results and discussion.

2. Experimental Setup and Procedure

The experiments were conducted with the bearings run-to-failure test data to validate the effectiveness of the proposed model for predicting the remaining life of the bearing. The entire experimental set-up used for data acquit ion is shown in Figure 1. This setup consists of a bearing, accelerometer, motor, DAQ card and lab VIEW software loaded into a computer to acquire the vibration signals.

The bearings are mounted on a shaft with housing and in-turn connected to a variable speed motor. Few major parameters like speed, load, temperature, lubrication, etc., are simulated at real time conditions. The brand new bearing (ball bearing 6205) are made to run to failure at rated speeds (1400 RPM) and rated axial loads (0.2 kN). The vibration signals from the bearings are collected through the accelerometers placed on the housings and acquired via DAQ card. These signals were processed in the Labview. The experiment was run in real time conditions matching to the said bearing application. Adequate measures were taken to avoid bearing faults or damages due to fitments, misalignments etc. The experiments on each bearing were conducted till the bearing fails naturally. For all the bearing monitoring data, vibration signal is more effective and suitable for reflecting bearing running condition. The amplitude of the vibration signals is monitored on a daily basis for

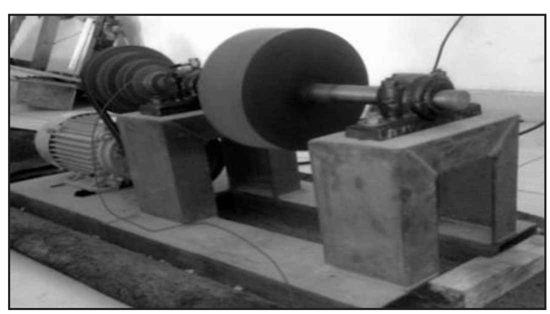
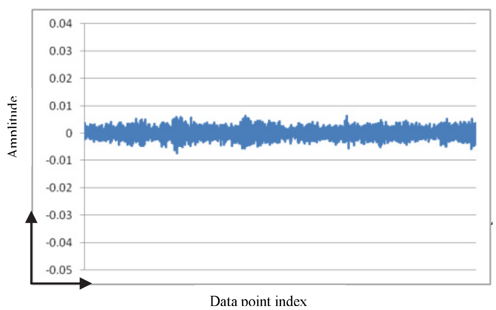
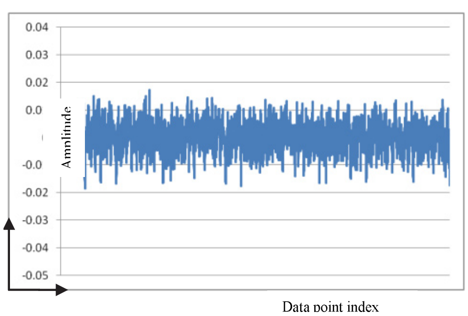


Figure 1. Experimental set-up.

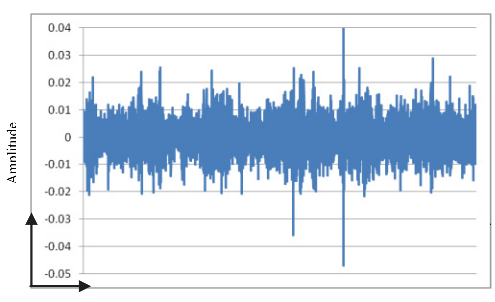
estimating the current health condition of the bearings. The vibration is small and smooth when the bearing is under normal condition. And the occurrence of bearing defect can cause fluctuation of vibration. During the process of degradation, the amplitude of vibration increases obviously as shown in the Figure 2. Thus, vibration signal becomes the convenient variable to estimate the remaining useful life of the bearings. Lifetime of bearing was categorized into five different stages viz, stage 1, stage 2, stage 3, stage 4 and stage 5 respectively. Signals acquired from a new bearing were placed in stage 1. Stage 2 and 3 includes the signals that are extracted from the bearing after 1000 hours and 1250 hours respectively. Stage 4 includes the signals that are extracted after 1500 hours of running the bearing at rated load and speed conditions. Signals from damaged bearing place in stage 5. The variations in signals, thus acquired for the bearings in various



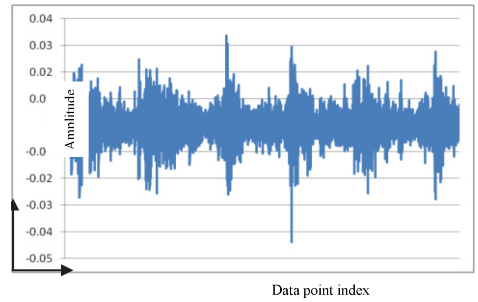
(a)



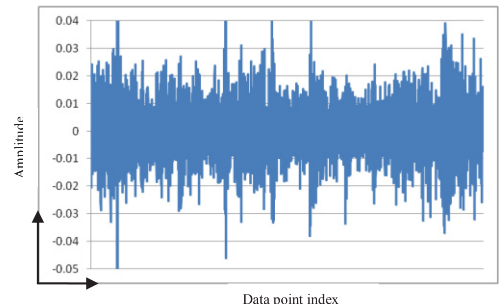
(b)



(c)



(d)



(e)

Figure 2. Vibration Signals acquired from experiments. **(a)** Stage-1: Vibration signals at start. **(b)** Satge-2:Vibration signals@1000Hrs. **(c)** Satge-3:Vibration signals@1250Hrs. **(d)** Satge-4:Vibration signals@1500Hrs. **(e)** Satge-5:Vibration signals@1800Hrs.

stages is shown in Figure 2 (a), 2 (b), 2 (c), 2 (d) and 2 (e) respectively.

3. Feature Description

Descriptive statistical parameters such as mean, median, mode, sample variance, kurtosis, skewness, standard error, standard deviation, minimum, maximum, sum, and range were computed to serve as features. They are named as 'statistical features' here[19,20]. Brief descriptions about the extracted features are given below.

3.1 Standard Error

Standard error is a measure of the amount of error in the prediction of y for an individual x in the regression, where x and y are the sample means and 'n' is the sample size.

Standard error of the predicted,

$$Y = \sqrt{\frac{1}{n-2}\left[\sum\left(y-\bar{y}\right)^2 - \frac{\left[\sum\left(x-\bar{x}\right)\left(y-\bar{y}\right)\right]^2}{\left(x-\bar{x}\right)^2}\right]} \qquad (1)$$

3.2 Standard Deviation

This is a measure of the effective energy or power content of the vibration signal. The following formula was used for computation of standard deviation.

$$\text{Standard Deviation} = \sqrt{\frac{\sum x^2 - \left(\sum x\right)^2}{n(n-1)}} \tag{2}$$

3.3 Sample Variance

It is the variance of the signal points and the following formula was used for computation of sample variance.

$$\text{Sample Variance} = \frac{\sum x^2 - \left(\sum x\right)^2}{n(n-1)} \tag{3}$$

3.4 Kurtosis

Kurtosis indicates the flatness or the spikiness of the signal. Its value is very low for normal condition of the bearing and high for the faulty condition of the bearing a due to the spiky nature of the signal.

$$\text{Kurtosis} = \left\{\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum\left(\frac{x_i-\bar{x}}{s}\right)^4\right\} - \frac{3(n-1)^2}{(n-2)(n-3)} \tag{4}$$

Where 's' is the sample standard deviation.

3.5 Skewness

Skewness characterizes the degree of asymmetry of a distribution around its mean. The following formula was used for computation of skewness.

$$\text{Skewness} = \frac{n}{n-1}\sum\left(\frac{x_i-\bar{x}}{s}\right)^3 \tag{5}$$

3.6 Range

It refers to the difference between maximum and minimum signal point values for a given signal.

3.7 Minimum Value

It refers to the minimum signal point value in a given signal. Therefore, it can be used to detect the faulty signal condition.

3.8 Maximum Value

It refers to the maximum signal point value in a given signal.

3.9 Sum

It is the sum of all feature values for each sample.

4. Classifier: Random Forest

Random Forest (RF) tool is one of the widely used classifiers in data mining techniques due to its accuracy among other learning algorithm[21]. It is more efficient when working with large database. RF is a combination of many decision trees. Decision trees are grown by a binary partitioning method to make easy interpretation.

The tree should be grown with the following conditions:

- Number of training sets be '*N*' and variables in the classifier be '*M*'.
- Let '*m*' be the input variable which is used to find the decision at the node, with $m<M$.
- Training case '*N*' is replaced by '*n*' times and also choose '*m*' variable random out of the '*M*' at each node of the tree. Best split is used at the nodes.
- The value '*m*' should be maintained constantly while growing the forest.
- Each tree is grown to maximum extent.

The error rate of the forest depends on the correlation of the two trees and strength of the individual tree. Generalization error is given by the formula,

$$\text{Generalization error} \le \frac{\bar{\rho}\left(1-s^2\right)}{s^2} \tag{6}$$

Where,

ρ is an average correlation among the trees.
s is a measure of strength of tree classifier.

5. Results and Discussions

The vibration signals from the bearings were collected at rated load and speed conditions[22]. A brand new bearing was taken for the experiment and signals were continuously taken on a daily basis till it fails naturally. The paper mainly focuses on building predictive model to assess the bearing lifetime. Statistical features were extracted and feature description is presented in section 4. It can be found using decision tree that all 12 statistical feature may not contribute equally towards the classification accuracy. Hence, parameters were optimized and only best contributing features (standard deviation, maximum, sum, skewness, and kurtosis) were selected for effective performance of the built model. The experiment setup was categorized into 5 different stages. Stage 1 which includes signals from new bearing. Stage 2 and 3 include the signals that are extracted after 1000 and 1250 hours of running the bearing at rated load and speed conditions respectively. Stage 4 includes the signals extracted at 1500 hours of running and damaged bearing signals were placed at stage 5.

5.1 Building and Testing of Predictive Model

The predictive model built using random forest was varied with respect to a number of seed and the number of trees. Figure 3 illustrate the classification accuracy with respect to seed. The seed is varied at intervals of '1' and the values were noted down and plotted in the graph (Figure 3). It is evident from the graph that when the number of seed is 4 the classification accuracy was maximum (95.64%).

Figure 4 illustrate the effect in a number of trees against the classification accuracy of the bearing. Here, the value is increased in the order of '10'. The values are noted down and the corresponding classification accuracy can be found (Figure 4). It is known from the graph, when the number of trees is '80' the classification accuracy is the maximum (95.52%).

The detailed accuracy and confusion matrix of the built is presented in Table 1 and Table 2 respectively. Out

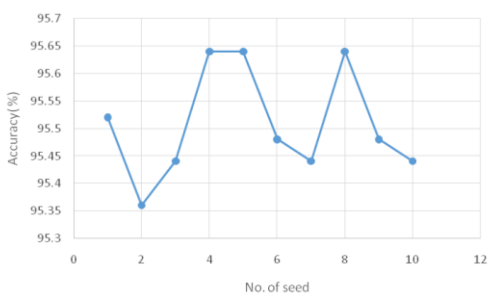


Figure 3. Effect of number of seeds on classification accuracy.

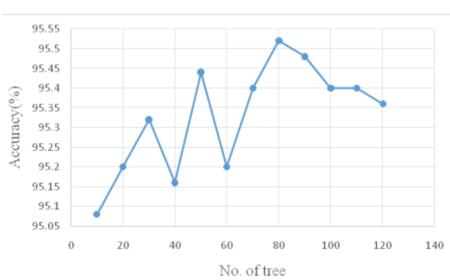


Figure 4. Effect in number of trees on classification accuracy.

Table 1. Detailed accuracy by class

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.996	0	1	0.996	0.998	1	Stage-1
	0.896	0.028	0.889	0.896	0.892	0.99	Stage-2
	0.89	0.027	0.894	0.89	0.892	0.991	Stage-3
	1	0	1	1	1	1	Stage-4
	1	0	1	1	1	1	Stage-5
Weighted average	0.956	0.011	0.956	0.956	0.956	0.996	

Table 2. Confusion matrix

a	b	c	d	e	Classified as
498	1	1	0	0	a = Stage-1
0	448	52	0	0	b = Stage-2
0	55	445	0	0	c = Stage-3
0	0	0	500	0	d = Stage-4
0	0	0	0	500	e = Stage-5

of many parameters in detailed accuracy table, TP rate and FP rate are important. TP rate denotes True Positive rate while the FP rate denotes the False Positive rate. Generally the TP rate should be close to 1 and the FP rate should be near to 0. In the present study TP rate and FP rate were found in 0.95 and 0.011 which can be accepted in many practical applications.

Confusion matrix can be found in Table 2. The interpretation of confusion matrix is as follows:

- The diagonal elements in the confusion matrix indicate the correctly classified instances.
- In the first row, the first element shows the number of data points that belong to 'stage -1' is correctly classified as 'stage -1'.
- In the first row, second and third element belongs to 'stage 1' was misclassified as 'stage 2' and 'stage 3' respectively.

Similarly, in all the rows, the diagonal elements represent correctly classified instances and the non-diagonal elements were interpreted as misclassified instances. In the second row, out of 500 instances 448 instances were correctly classified and 52 instances were misclassified as 'stage 3' and in row 3, 55 instances were misclassified as 'stage 2'. There were no misclassifications in stage 4 and stage 5. It should be noted that the overall error percentage is less than 5%, which is acceptable for in practical applications.

6. Conclusion

The research paper presented a predictive model based on a classification approach to assess the remaining useful time of bearing. In the present study, the vibration signals were acquired from the bearing continuously till it fails naturally. Statistical features were extracted and best contributing features are selected using decision tree. With the selected feature, the predictive model was built. The classification accuracy was found to be 95.64% and error percentage is less than 5% which can be accepted practically. Hence, the proposed model can be used for predicting the remaining lifetime of bearings successfully.

7. References

1. Palmgren A, vonD L, Kugellagern. The service life of ball bearings. Zeitschrift des Vereines DeutscherIngenieure. 1924; 68(4):339–41.
2. Anon. Load ratings and fatigue life for ball bearings. The Anti-Friction Bearing Manufacturers Association, ANSI/AFBMA; Washington, DC. 1990.
3. Ramroop G, LiuK, GuF, Paya BS, Ball AD. Airborne Acoustic Condition Monitoring. 2003; Available from: www.maintenenceengineering.com
4. Heng RBW, Nor MJM. Statistical analysis of sound and vibration signals for monitoring rolling element bearing condition, Applied Acoustics. 1997; 53(1–3):211–66.
5. Shibata K, Takahashi A, Shirai T. Fault diagnosis of rotating machinery through visualization of sound signals. Mechanical Systems and Signal Processing. 2000; 14(2):229–41.
6. Gebraeel N, Lawley M, Liu R, Parmeshwaran V. Residual life predictions from vibration-based degradation signals: A neural network approach. IEEE Transactions on Industrial Electronics. 2004; 51(5):694–700.
7. Liu T, Ordukhani F, Dipak J. Monitoring and diagnosis of roller bearing conditions using neural networks and soft computing. International Journal of Knowledge-based and Intelligent Engineering Systems. 2005; 9(2):149–57.
8. Qui H, Lee J, YuG. Robust performance degradation assessment methods for enhanced rolling elements bearing prognostics. Advanced Engineering Informatics. 2003 Jul-Oct; 17(3-4):127–40.
9. Zhao Y, Zhang G, Du J, Wang G, Vachsevanos G. Development of distributed bearing health monitoring and assessing system. 8th International Conference on Control, Automation, Robotics and Vision; 2004. p. 474–8.
10. Muralidharan V, Sugumaran V. A comparative study of Naive Bayes classifier and Bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis. Applied Soft Computing. 2012 Aug; 12(8):2023–9.
11. DaSilva V, Fujimoto RY, Padovese LR. Rolling bearing fault diagnostic system using fuzzy logic. 10th International conference on Fuzzy Logic; 2001. p. 816–9.
12. Artes M, Castillo DL, Perez J. Failure prevention and diagnosis in machine elements using cluster. Proceeding of the 10th International Congress on Sound and Vibration; 2003. p. 119– 1203.
13. Yu WK, Harris TA. A new stress-based fatigue life model for ball bearings. Tribology Transactions. 2001; 44(p1):11–8.
14. Li Y, Billington S, Zhang C, Kurfess T, Danyluk S, Liang S. Dynamic prognostic prediction of defect propagation on rolling element bearings. Tribology Transactions. 1999; 42(2):385–92.
15. Qiu J, Seth BB, Liang S, Zhang C. Damage mechanics approach for bearing lifetime prognostics. Journal of Mechanical Systems and Signal Processing. 2002; 16(5):817–29.
16. Li Y, Kurfess T, Liang S. Stochastic prognostics for rolling element bearings. Journal of Mechanical Systems and Signal Processing. 2000; 14(5):747–62.
17. Shao Y, Nezu K. Prognosis of remaining bearing life using neural networks. Proc Inst Mech Eng J Syst Control Eng. 2000; 214(3):217–30.
18. Gebraeel N, Lawley M, Liu R, Parmeshwaran V. Residual life predictions from vibration-based degradation signals: A neural network approach. IEEE Transactions on Industrial Electronics. 2004; 51(3):694–700.
19. Jagadeeswaran R, Sugumaran V. Comparative study of decision tree classifier and best first tree classifier for fault diagnosis of automobile hydraulic brake system using statistical features, Measurement. 2013; 46(9):3247–60.
20. Muralidharan V, Sugumaran V. Feature extraction using wavelets and classification through decision tree algorithm for fault diagnosis of mono-block centrifugal pump. Measurement. 2012; 46(1):353–9.
21. Sasikala S, Bharathidason S, Venkateswaran CJ. Improving classification accuracy based on random forest model through weighted sampling for noisy data with linear decision boundary. Indian Journal of Science and Technology. 2015 Apr; 8(S8). DOI: 10.17485/ijst/2015/v8iS8/71714.
22. Mazurkow A. The analysis of the influence of structural parameters on static and dynamic properties of sliding bearings with a floating ring bearing. Indian Journal of Science and Technology. 2010 Jun; 3(6). DOI: 10.17485/ijst/2010/v3i6/29767.