Fault Diagnosis of Helical Gear Box Using Decision Tree and Best-First Tree

1M. Amarnath, 2Deepak Jain, 3V. Sugumaran and 4Hemantha Kumar

1Department of Mechanical Engg., Indian Institute of Information Technology Design and Manufacturing Jabalpur, Jabalpur, Madhyapradesh, India,

2,3SMBS, VIT University, Chennai Campus, Vandalur-Kelambakam Road, Chennai,

4Department of Mechanical Engineering, National Institute of Technology, Surathkal, Karnataka, Mangalore, India,

ABSTRACT

Gear is one of critical transmission elements, found its wide applications in small, medium and large machineries. It is well proved that vibration signals acquired from a rotating machine comprise of the dynamic information about the health condition of the rotating machine. This paper uses vibration signals acquired from machinery comprising of gears in good and simulated faulty conditions for the purpose of fault diagnosis by machine learning approach. In this paper, a vibration based on machine learning approach is presented for helical gear box as it plays critical role in the industries. This approach has mainly three steps namely feature extraction, feature selection and classification. Here statistical analysis was used for feature extraction and decision tree and best-first decision tree for classification was taken and compared. The classification accuracies for different conditions were calculated and compared to find the best classifier for the fault diagnosis of the helical gear box.

Keywords: Gear fault diagnosis, Fault diagnosis, Statistical analysis, Decision tree, Best-first Decision tree

1. INTRODUCTION

Helical gear box condition monitoring has received significant attention from many years. The distinctive failure mode of helical gear box is localized defects, which occur if a sizable piece of material on the contact surface is shifted during operation, usually by fatigue cracking under cyclic contact stressing. Hence failure alarm from helical gear box is frequently established on the detection of the localized defects in early stage. If the helical gear box is operating under various speeds and loads, it’s not possible to measure or estimate the inclemency of localized faults, hence certain physical parameters such as vibration, sound, acoustic emission and wear debris have been considered in detection and diagnosis of inchoate faults. Sound and vibrations generated by rotating machinery often mask the features of fault-related signals generated by the machine elements like gears, bearings and cams (Lai Wuxiang et al. [1], Zvokelj et al. [2], Li and Ma [3] and Tinta et al. [4]). Then again, it is known that local faults in helical gear box induce impacts, as a result of which transient excitations might be seen in vibration and sound signals. Villa et al. [5] presented findings which deals with the detection of different mechanical faults (unbalance and misalignment) under a broad range of working conditions of speed and load. The conditions tested in a test bench are alike to the ones that can be
found in dissimilar kinds of machines like for example wind turbines. The authors demonstrate how to take advantage of the information on vibrations from the mechanical system under study in a broad range of load and speed conditions. Utilizing such information the prognosis and detection of faults is quicker and more reliable than the one obtained from an analysis over a limited range of working conditions (e.g. nominal). Due to the presence of growing local faults, vibration and sound signals from helical gear box have non-stationary characteristics; therefore analyses of the above signals in faulty operating conditions become hard. Under such conditions, ceremonious application of measurement of statistical parameters may not be practicable (Oguamanam et al. [6], Benko et al. [7] and Benko et al. [8]). Loutas et al. [9] conducted analyses on lab-scale, single stage, gearbox using different non-destructive inspection methodologies with advanced signal processing techniques. Ramroop et al. [10] reported a comparative study of conventional vibration and acoustic monitoring techniques to diagnose defects in industrial multi stage gearbox, operating under healthy and faulty conditions. Tekiner and Yesilyurt [11] carried out experiments to find best suitable machining conditions using statistical values of sound signals. Their work highlighted the influence of sound signal to analyze the machining process parameters viz. flank wear, surface roughness, chip morphology and built-up edge formation. Byrtus and Zeman [13] presented an original method of the mathematical modeling of gear drive nonlinear vibrations. Their proposed model includes nonlinearities caused by gear mesh interruption, parametric gearing excitation caused by time-varying meshing stiffness and nonlinear contact forces acting between journals of the rolling-element bearings and the outer housing. The nonlinear model is then used for investigation of gear drive vibration, especially for detection of nonlinear phenomena like impact motions, bifurcation of solution and chaotic motions in case of small static load and in resonant states. The theoretical method is used for investigation of two-stage gearbox nonlinear vibration. Amarnath et al. [14] presented the results of experimental investigations carried out to assess wear in spur gears of back-to-back gearbox under accelerated test conditions. The studies considered the estimation of operating conditions such as film thickness and their effects on the fault growth on teeth surface. Modal testing experiments have been carried out on the same gear starting from healthy to worn out conditions to quantify wear damage. The results provide a good understanding of dependent roles of gearbox operating conditions and vibration parameters as measures for effective assessment of wear in spur gears. Saravanan et al. [15] presented the effectiveness of wavelet-based features for fault diagnosis of a helical gear box using artificial neural network (ANN) and proximal support vector machines (PSVM). The statistical feature vectors from Morlet wavelet coefficients are classified using J48 algorithm and the predominant features were given as input for training and testing ANN and PSVM and their relative accuracy in classifying the faults in the bevel gear box was compared. The Naïve Bayesian classifier is very simple and efficient and highly sensitive to feature selection, so the research of feature selection especially for it is significant. Naïve Bayes and Bayes net algorithms were effectively used for tool condition monitoring as well [16].

From the literature one can understand that many classification algorithms have been used for classifying the faults in helical gear box and other rotating members. In order to say strongly that a particular algorithm is better compared to other algorithms a detailed comparative study needs to be done. Therefore, this paper mainly deals with the performances of Decision Tree and Best-first Decision Tree algorithms in classifying gear faults.
2. EXPERIMENTAL SETUP AND PROCEDURE

Fig.1 shows the experimental setup. The setup consists of a 5 HP two stage helical gearbox. The gearbox is driven by a 5.5 HP, 3-phase induction motor with a rated speed of 1440 rpm. The speed of the motor is controlled by an inverter drive and for the present study the motor is operated at 80 rpm. In other words the speed of the first stage of the gearbox is 80 rpm. With a step-up ratio of 1:15, the speed of the pinion shaft in the second stage of the gear box is 1200 rpm. Table 1 summarizes the specifications of the test rig.

<table>
<thead>
<tr>
<th>Specifications of helical gearbox</th>
<th>First stage</th>
<th>Second stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of teeth</td>
<td>44/13</td>
<td>73/16</td>
</tr>
<tr>
<td>Pitch circle diameter (mm)</td>
<td>198 /65</td>
<td>202 /48</td>
</tr>
<tr>
<td>Pressure angle (°)</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Helix angle</td>
<td>20</td>
<td>15</td>
</tr>
<tr>
<td>Modules</td>
<td>4.5 / 5</td>
<td>2.75 / 3</td>
</tr>
<tr>
<td>Speed of shafts</td>
<td>80 rpm (input)</td>
<td>1200 rpm (output)</td>
</tr>
<tr>
<td>Mesh frequency</td>
<td>59 Hz</td>
<td>320 Hz</td>
</tr>
<tr>
<td>Step-up ratio</td>
<td>1:15</td>
<td></td>
</tr>
<tr>
<td>Rated power</td>
<td>5 HP</td>
<td></td>
</tr>
<tr>
<td>Power Transmitted</td>
<td>2.6 HP</td>
<td></td>
</tr>
</tbody>
</table>

The pinion is connected to a D.C motor (which is used as generator) to generate 2 kW power, which is dissipated in a resistor bank. Therefore, the actual load on the gearbox is only 2.6 HP which is 52% of its rated power 5 HP. In industrial environment utilization of load varies from 50% to 100%. In the case of traditional dynamometer, additional torsional vibrations can occur due to torque fluctuations. This is avoided in this case by using D.C motor and resistor bank.

Tyre couplings are fitted between the electrical machines and gear box so that backlash in the system can be restricted to the gears. The motor, gear box and generator are mounted on I-beams, which are anchored to a massive foundation. Vibration signals are measured using a B&K accelerometer which is installed close to the test bearing. Signals are sampled at a sampling frequency of 8.2 kHz. The experimental setup with equipment and sensors is shown in Fig. 2.
Overhaul time of a new gear box is more than one year. It is very difficult to study the fault detection procedures without seeded fault trials. Local faults in a gear box can be classified into three categories. (a) Surface wear spalling (b) cracked tooth and (c) loss of a part of tooth due to breakage of tooth at root or at a point on working tip (broken tooth or chipped tooth). There are different methods to simulate faults in gears viz. electric discharge machining (EDM), grinding, adding iron particles in gearbox lubricant and over loading the gear box i.e. accelerated test condition. The simplest approach is partial tooth removal. This simulates the partial tooth break, which is common in many industrial applications (Staszewski et al. [17], Yesilyurt [18], Yesilyurt [19] and Loutridis [20]).

3. FEATURE EXTRACTION

Statistical analysis of vibration signals concedes different parameters. The statistical parameters used for this study are mean, standard error, median, standard deviation, sample variance, kurtosis, skewness, range, minimum, maximum and sum. These features are extracted from the vibration signals. They are named as ‘statistical features’. Brief descriptions of the extracted features are given in the following Table 2.

Table 2- Statistical Features

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Name of the Feature</th>
<th>Feature Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Standard error</td>
<td>Standard error of the predicted, $Y = \sqrt{\frac{1}{n-2} \sum (y - \bar{y})^2 - \left(\frac{\sum (x - \bar{x})(y - \bar{y})}{x - \bar{x}}\right)^2}$</td>
</tr>
<tr>
<td>2</td>
<td>Standard deviation</td>
<td>Standard Deviation = $\sqrt{\frac{\sum x^2 - \left(\frac{\sum x}{n}\right)^2}{n(n-1)}}$</td>
</tr>
<tr>
<td>3</td>
<td>Sample variance</td>
<td>Sample Variance = $\frac{\sum x^2 - \left(\frac{\sum x}{n}\right)^2}{n(n-1)}$</td>
</tr>
</tbody>
</table>
4. **CLASSIFIER**

Decision tree is a tree established on knowledge methodology applied to interpret classification rules [21]. A standard tree got from J48 comprises of a number of branches, one root, a number of nodes and a number of leaves (See Fig.4). One branch of which is a chain of nodes from root to a leaf and each node have one attribute. An attribute in a tree gives the information of the importance of the associated attribute. The procedure of making the Decision Tree and using the same for feature selection is explained in detail by Sugumaran et al [21].

4.1 **Gini Index**

Gini index is another criterion to evaluate the impurity of a node, which is applied in the CART system (Breiman et al [22]). If \( p_i \) is probability that the instance is in class \( i \) and \( p_j \) stand for the probability that an instance is in class \( j \), the Gini index is given in form (Breiman et al [22]).

\[
gini(p_1, p_2, \ldots, p_n) = \sum_{j \neq i} p_i p_j
\]

(1)

The Gini index has an interesting interpretation in terms of sample variances (Breiman et al [22]). When all the instances at a node that are in class \( j \) are allotted the value 1 and the other instances are allotted the value 0, the sample variance for these values is \( p_j (1 - p_j) \), i.e. \( p_j - p_j^2 \). Restating this for all classes and adding the variances as the sum of the \( p_j \) over all classes is 1, it can be written that the Gini index can be presented as

\[
gini(p_1, p_2, \ldots, p_n) = \sum_{j} (p_j (1 - p_j)) = 1 - \sum_j p_j^2
\]

(2)

If only two classes are used, the Gini index can be modified to \( 2p_1p_2 \). Figure 3 tells the relationship between the Gini index and the class probability of the first class in a two-class problem, accompanied by the relationship between the information value and the same class probability. From the figure, we can view that in both cases, the impurity is maximum if the two classes has equal probability (i.e. 0.5 for each class). And if any of class probability is 1, the impurity is 0. This means the node is absolutely pure.
5. RESULTS AND DISCUSSION

The vibration signals were recorded for normal and abnormal conditions of helical gear box. Totally 420 samples were collected; out of which 60 sample were from no load condition. For 100% load with 10%, 20%, 30%, 40%, 80% and 100% fault, 60 samples from each condition were collected. The statistical parameters explained in Section 3 are treated as features and used as input to the algorithm. The corresponding status or condition (10%, 20%, 30%, 40%, 80%, 100%, expt no load) of the classified data will be the required output of the algorithm. This input and corresponding output together forms the dataset.

5.1 Effect of number of features and Feature selection

Totally fourteen descriptive statistical features were extracted from vibration signals. All of them may not be significant for the classification purpose. One cannot say before in hand which are the features will be helpful for the purpose of classification while using machine learning algorithms. As more number of irrelevant features actually may reduce the performance of the classification algorithm. Also, they increases the computational resources required. Researchers have to extract all descriptive statistical features and then select the good ones. Here, decision tree was used for feature selection. In decision tree, the feature that occurs first will be the root node and the same will be the best feature for classification. The other features in the tree are in the order of importance. In decision tree, only seven features were present namely, standard error, maximum, minimum, range, mean, median and kurtosis in the order of importance. This situation in general means only seven features are enough to classify the gear box conditions. However, one needs to confirm the same to be on the safer side. For this purpose, the effect of features on classification accuracy study was carried out (Sugumaran et al [21]). Initially, only first best feature alone was used with decision tree and found the classification accuracy. This happens to be mean as it is in the top node of the decision tree. Then, top two best features were used with decision tree and found the classification accuracy (mean and standard error).
5.2 Decision Tree

The dataset is used with decision tree algorithm for the purpose of feature selection and classification. The generated decision tree is shown in Fig. 4. The rectangles represent classes (condition of the helical gear box). In rectangle the information about the condition is given using abbreviations e.g. ‘expt no load’. Then within parenthesis, there are two numbers separated by a slash or one number. The first number (in case of two numbers) or the only number represents the number of data points that support the decision. Meaning, if one follows the rule (as described above ‘if-then’ rules) how many data points will be correctly classified is given as first number. The second number (after slash) is optional and it represents the number of data points that is against the rule followed. Meaning, if one follows a rule, how many data points will be incorrectly classified is given as second number.

The detailed class-wise accuracy of the J48 algorithm is presented in Fig. 5. Out of the terms used in Fig. 8, ‘TP rate’ and ‘FP rate’ are very important. The ‘TP rate’ stands for true positive and its value should be close to ‘1’ for better classification accuracy. The ‘FP rate’ stands for false positive and its value should be close to ‘0’ for better classification accuracy. In the study, one can appreciate the closeness of ‘TP rate’ to ‘1’ and ‘FP rate’ to ‘0’. The both values confirm that the built model is good one.
The classification accuracy of the decision tree algorithm is presented in Fig.9. The interpretation of the confusion matrix is as follows:

- The diagonal elements in the confusion matrix (Refer Fig.6) show the number of correctly classified instances.
- In the first row, the first element shows number of data points that belong to ‘10.0% fault’ class and classified by the classifier as ‘10.0 fault’.
- In the first row, the fifth element shows the number of data points belonging to ‘10.0% fault’ class but misclassified as ‘80% fault’.
- In the first row, the sixth element shows the number of ‘10% fault’ data points misclassified as ‘100.0% fault’.
- In the first row, the seventh element shows the number of ‘10% fault’ data points misclassified as expt no load and so on.

In the present study, the faulty conditions are clearly distinguishable from 10% fault condition by the algorithm. Therefore, one can note ‘0’ in first column third and fourth elements which correspond to misclassification of faulty conditions. However, there are misclassifications in other conditions. They are given in non-diagonal elements. Here, out of 420 data points, 65 data points were misclassified by the algorithm. Actually, this is an error of about 15.5%, which is acceptable for many practical applications.
Summary of stratified cross validation is given below.

<table>
<thead>
<tr>
<th>Total Number of Instances</th>
<th>420</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctly Classified Instances</td>
<td>355</td>
</tr>
<tr>
<td>Incorrectly Classified Instances</td>
<td>65</td>
</tr>
</tbody>
</table>

5.3 Best First Tree Algorithm

From the decision tree, seven features that are contributed for classification were only selected for training and testing. The selected features were trained and tested by using Best first tree algorithm with post pruning processes.

The detailed class-wise accuracy of the Best first tree is presented in Fig.7. Misclassification details for both the conditions are presented as confusion matrix in Fig.8. Out of the terms used in Fig.8, ‘TP rate’ and ‘FP rate’ are very important. The ‘TP rate’ stands for true positive and its value should be close to ‘1’ for better classification accuracy. The ‘FP rate’ stands for false positive and its value should be close to ‘0’ for better classification accuracy. In the study, from Fig.8, one can appreciate the closeness of ‘TP rate’ to ‘1’ and ‘FP rate’ to ‘0’. The both values confirm that the built model is good one. The classification accuracy of the best first tree algorithm is presented in Fig.7.

![Table](image)

Fig.7 Detailed accuracy by class - Best first tree

In the present study, the faulty conditions are clearly distinguishable from 10% fault condition by the algorithm. Therefore, one can note ‘0’ in first column third and fourth elements which correspond to misclassification of faulty conditions. However, there are misclassifications in other conditions. They are given in non-diagonal elements. Here, out of 420 data points, 62 data points were misclassified by the algorithm. Actually, this is an error of about 14.7 %, which is acceptable for many practical applications.


The Summary of stratified cross validation is given below.

<table>
<thead>
<tr>
<th>Total Number of Instances</th>
<th>420</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctly Classified Instances</td>
<td>358</td>
</tr>
<tr>
<td>Incorrectly Classified Instances</td>
<td>62</td>
</tr>
</tbody>
</table>

Fig. 9 illustrates the comparison of classification efficiency of J48 Decision tree and Best first decision tree algorithm with variation in number of objects.

6. CONCLUSION

Condition monitoring of a helical gear box was carried out using vibration signals and the statistical features were extracted. They were classified using decision tree and best first tree classifiers and the superior feature – classifier combination was found. Dimensionality reduction was carried using decision tree. The classification accuracy of statistical feature using decision tree was found to be 84.52% while with best first tree it was found to be 85.23%. The statistical feature with best first tree combination yields the highest accuracy as per the conditions in this study. Although the difference in classification accuracy between the best first tree and decision tree may be marginal, the kind of features makes a huge impact on the study results.
7. REFERENCES


