Expert Systems with Applications 38 (2011) 3819–3826

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Multi component fault diagnosis of rotational mechanical system based on decision tree and support vector machine

M. Saimurugan [a], K.I. Ramachandran [a], V. Sugumaran [b,*], N.R. Sakthivel [a]

[a] Department of Mechanical Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, Tamilnadu, India
[b] Department of Mechanical Engineering, SRM University, Kattankulathur, Kancheepuram, Tamilnadu, India

ARTICLE INFO

Keywords:
Fault diagnosis
Shaft and bearings
Decision tree
Support vector machine

ABSTRACT

The shaft and bearing are the most critical components in rotating machinery. Majority of problems arise from faulty bearings in turn affect the shaft. The vibration signals are widely used to determine the condition of machine elements. The vibration signals are used to extract the features to identify the status of a machine. This paper presents the use of c-SVC and nu-SVC models of support vector machine (SVM) with four kernel functions for classification of faults using statistical features extracted from vibration signals under good and faulty conditions of rotational mechanical system. Decision tree algorithm was used to select the prominent features. These features were given as inputs for training and testing the c-SVC and nu-SVC model of SVM and their fault classification accuracies were compared.

1. Introduction

In last two decades, due to the increase in production capabilities of modern manufacturing systems, plants are expected to run continuously for extended hours. As a result, unexpected downtime due to machinery failure has become more costly than ever before. Therefore, condition monitoring is gaining importance in industry because of the need to increase machine availability and health trending, to warn of impending failure. It is required to detect, identify and then classify different kinds of faults that can occur within a machine. Often several different kinds of sensors are employed at different positions to sense a variety of possible faults. Features are then computed to analyze the signals from all these sensors to assess the health of the machine or its components. Thus, the application of a condition monitoring-based maintenance policy can help to minimize unnecessary costs and delays caused by unscheduled repairs.

The majority of problems in rotational mechanical system arise from faulty bearings which in turn affect the shaft (Winder & Littmann, 1976). An air crash was caused by bearing failure resulting from damage to the cage supporting the bearing balls (Smalley, Baldwin, Mauney, & Millwater, 1996). The reports (Kalkat, Yildirim, & Uzmay, 2005; Sun, Chen, & Li, 2007) have clearly described the fault diagnosis of rotational mechanical system which focused on faulty shaft. Vibration-based monitoring techniques, both in the time and frequency domains, have been widely used for detection and diagnosis of bearing defects for several decades. A brief review of vibration monitoring techniques can be found in Cempel (1991). The report (Su & Lin, 1992) has extended a previous vibration model proposed to describe the bearing vibration caused by a single defect, giving a detailed insight into the analysis of vibration spectra.

Data mining is the process of discovering knowledge from large amount of data, which needs no prior domain knowledge. Data mining has many topics such as classification, clustering, association and prediction. Recently, classification problem is the research hotspot and SVM is one of the most widely used classification methods. SVM is used in many applications of machine learning because of its high accuracy and good generalization capabilities (Burgess, 1998). SVM is based on statistical learning theory. Recent work (Jack & Nandi, 2000) reports the use of SVM for on-line condition monitoring and its comparison with Artificial Neural Network (ANN). SVM classifies better than ANN because of the principle of risk minimization. Most recent SVM works are on fault diagnosis of roller bearing (Abbasion, Rafsanjani, Farshidianfar, & Irani, 2007; Sugumaran, Muralidharan, & Ramachandran, 2007; Sugumaran, Sabareesh, & Ramachandran, 2008; Yang, Yu, & Cheng, 2007a; Yang, Zhang, & Zhu, 2007b).

All features contribute to the classifications and there is a need for selecting the best features before classification. There are many techniques available for feature selection. The commonly used techniques for selection of features are decision tree (Samanta & Al-Baulshi, 2003) and principal component analysis (PCA) (Salido & Shuta, 2004). CART (Breiman, Friedman, & Olshen, 1984) and C4.5 (Quinlan, 1993) are the two best-known and widely used algorithms. In the present study, C4.5 decision tree is used for feature selection.

* Corresponding author. Tel.: +91 9486094486.
E-mail address: v_sugu@yahoo.com (V. Sugumaran).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.09.042

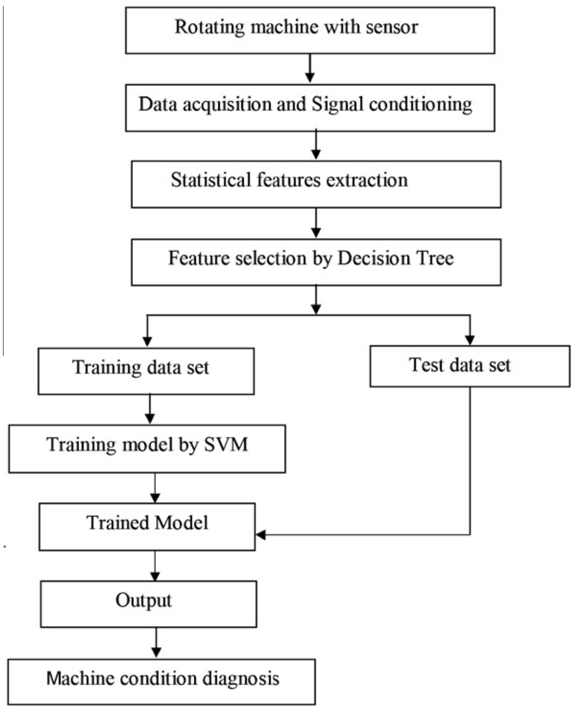

Fig. 1. Flow chart of fault diagnostic procedure.

Lot of work has been carried out on the fault diagnosis of shaft or bearing individually, and in the present study, both shaft and bearings are considered together. This paper addresses the use of decision tree and support vector machine for feature selection and classification respectively. Fault diagnostic procedure is shown in Fig. 1.

The main contribution of the paper is as follows: the vibration signal from a piezoelectric transducer is captured for 12 different conditions such as bent shaft, unbalanced shaft, bearings with inner race fault, etc. The statistical features were extracted and best features were selected using decision tree then it was classified successfully using c-SVC and nu-SVC model of SVM with four kernel functions such as linear function, three degree polynomial function, radial basis function (RBF) and sigmoid function. Then their fault classification accuracies are compared.

The rest of paper is organized as follows. Section 2 details about the experimental study which includes the experimental setup and procedure. Section 3 describes the statistical features extraction from the vibration signals. Section 4 discuss is about the decision tree algorithm. The support vector machine is explained in detail in Section 5. Section 6 analyzes the experimental results and compares their performance. The conclusions are presented in final section.

2. Experimental studies

The experimental setup comprising of the fault simulator with sensor and data acquisition and experimental procedure are described below.

2.1. Experimental setup

The experimental setup is shown in Fig. 2. A variable speed DC motor (0.5 hp) with speed up to 3000 rpm is the drive. A short shaft of 30 mm diameter is attached to the shaft of the motor through a flexible coupling. The flexible coupling is used to minimize effects of misalignment between the shafts and to transmit the power effectively from the motor. The shaft is supported at its ends through two ball bearings. The one, which is closer to the motor, is a new bearing (assumed to be free from defects). The bearing at the farther end is the bearing under test; provision is also made to change it easily. The sensor is to be mounted on top of the bearing housing. The selected area is made flat and smooth to ensure effective coupling. Provisions are there to mount three discs in the shaft, where the eccentric loading is possible.

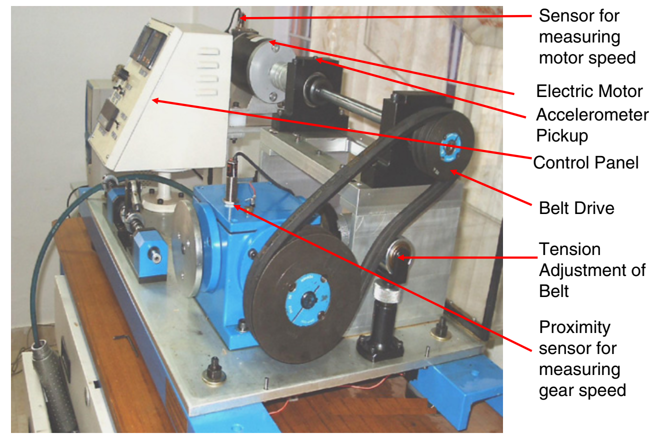

Fig. 2. Experimental setup.

A piezoelectric accelerometer (Dytran model) is mounted on the flat surface using direct adhesive mounting technique. The accelerometer is connected to the signal-conditioning unit (DACTRON FFT analyzer), where the signal goes through the charge amplifier and an Analogue-to-Digital converter (ADC). The vibration signal in digital form is fed to the computer through an USB port. The software RT Pro-series that accompanies the signal conditioning unit is used for recording the signals directly in the computer's secondary memory. The signal is then read from the memory and replayed and processed to extract different features.

2.2. Experimental procedure

In the present study, four SKF 6206 ball bearings were used. Out of four bearings two were new bearings and free from defects. In the other two ball bearings, defects were created using EDM in order to keep the size of the defect under control. The size of inner race defect is 0.552 mm wide and 0.782 mm deep and that of outer race defect is 0.625 mm wide and 0.974 mm deep. Fig. 3 show the Inner race fault and Outer race fault bearing. Before installing, each bearing was properly lubricated with grease.

The vibration signal from the piezoelectric pickup was taken after allowing initial running for sometime. The sampling

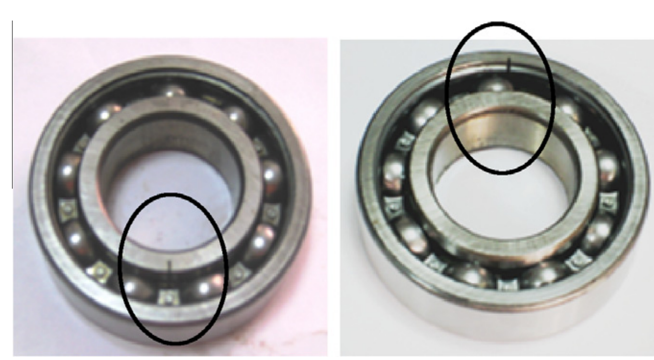

Fig. 3. Inner race fault and Outer race fault bearing.

frequency was 12000 Hz and sample length was 8192 for all speeds and all 12 conditions. The twelve different conditions and their notation used are good shaft good bearing(a1), unbalanced shaft good bearing(a2), good shaft with inner race fault (IRF) bearing(a3), unbalanced shaft with IRF bearing(a4), good shaft with outer race fault (ORF) bearing(a5), unbalanced shaft with ORF bearing(a6), shaft bent good bearing(a7), shaft bent with IRF bearing(a8), shaft bent with ORF bearing(a9), shaft bent with unbalancing mass and good bearing(a10), shaft bent with unbalancing mass and IRF bearing(a11), shaft bent with unbalancing mass and ORF bearing(a12). Many trials were taken at the speed of 500 rpm, 700 rpm, 900 rpm and 1100 rpm and the vibration signals were stored in the data file. Fig. 4 shows the time domain signals taken from different conditions. They show time domain plots of vibration acceleration of all twelve different conditions at 500 rpm.

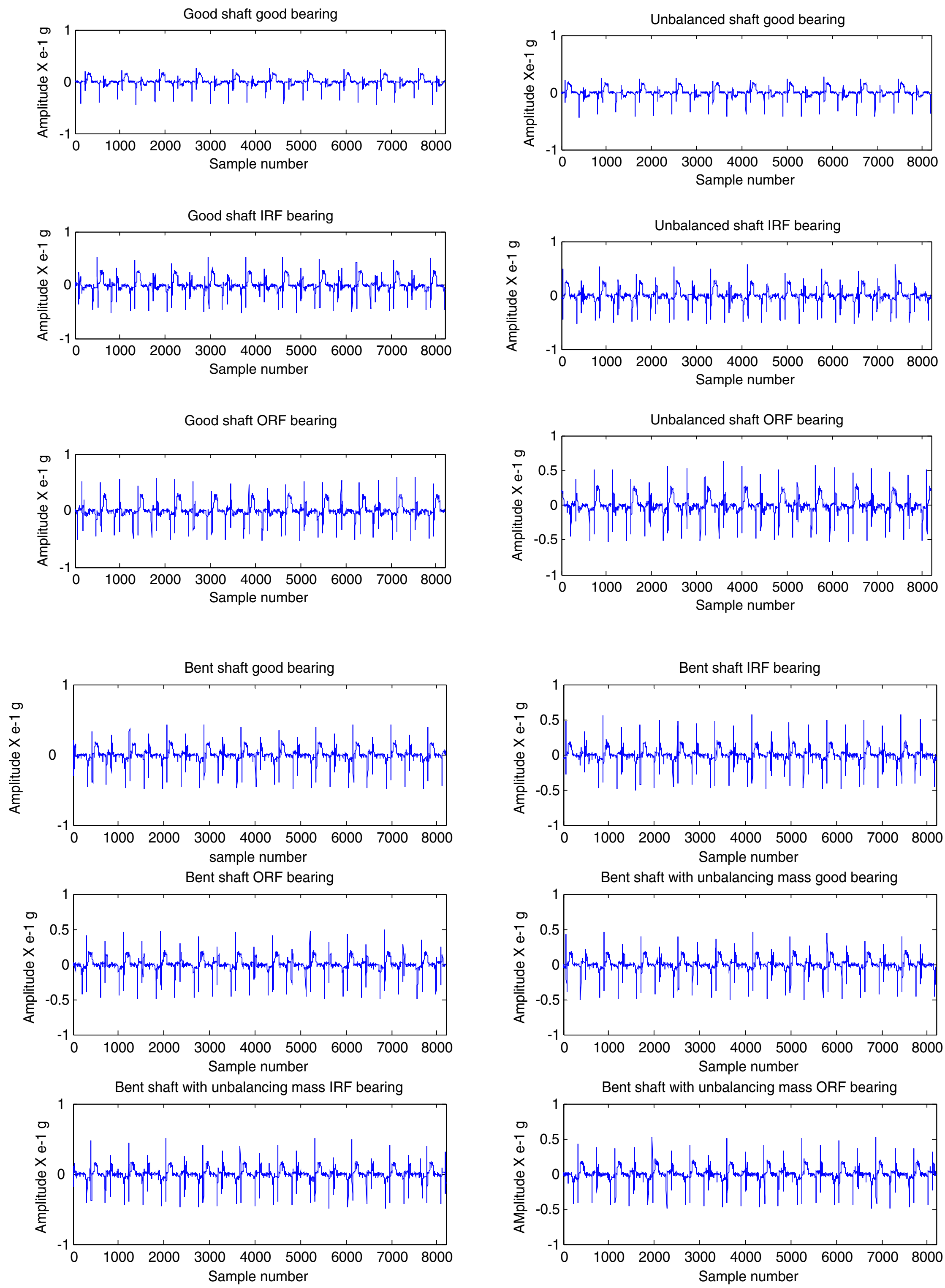

Fig. 4. Time domain plots of signals.

3. Feature extraction

Statistical analysis of vibration signals yields different parameters. Research work reported (Shi, Xu, & Xu, 1988) uses these in combinations to elicit information regarding the bearing faults. Such procedures use allied logic often based on physical considerations. In this study, a fairly wide set of the parameters were selected. They are mean, standard error, median, standard deviation, sample variance, kurtosis, skew ness, range, minimum, maximum and sum. These statistical features are explained below.

(a) 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

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

(b) 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.

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

(c) Sample variance: It is variance of the signal points and the following formula was used for computation of sample variance.

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

(d) Kurtosis: Kurtosis indicates the flatness or the spikiness of the signal. Its value is very low for normal condition of the pump and high for faulty condition of the pump 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.

(e) Skewness: Skewness characterises the degree of asymmetry of a distribution around its mean. The following formula was used for computation of skewness.

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

(f) Range: It refers to the difference in maximum and minimum signal point values for a given signal.

(g) Minimum value: It refers to the minimum signal point value in a given signal. As the pump parts (impeller, seal, bearing) get degraded, the vibration levels seem to go high. Therefore, it can be used to detect faulty pump condition.

(h) Maximum value: It refers to the maximum signal point value in a given signal.

(i) Sum: It is the sum of all signal point values in a given signal.

These features were extracted from vibration signals.

4. Decision tree (C4.5 Algorithm)

Decision tree represents the information in the signal as features, in the form of a tree. The classification is done through the decision tree with its leaves representing the different conditions of the engine. The sequential branching process ending up with the leaves here is based on conditional probabilities associated with individual features.

C4.5 algorithm introduced by Quinlan (1996) is one of the widely used algorithms to generate decision tree. Prior to the presentation of the specific role of C4.5 algorithm, the underlying theory is presented.

4.1. Building the decision tree

In the building phase, the training sample sets with discrete-valued attributes are recursively partitioned until all the records in a partition have the same class. The tree has a single root node for the entire training set. A new node is added to the decision tree for every partition. For a set of samples in a partition S, a test attribute X is selected for further partitioning the set into $S_1, S_2, S_3, \ldots, S_L$. For each new set $S_1, S_2, S_3, \ldots, S_L$ new nodes are created and these are added to the decision tree as children of the node for S. Further, the node for S is labeled with test X, and partitions $S_1, S_2, S_3, \ldots, S_L$ are recursively partitioned. When all the records in a partition have identical class label, further portioning is stopped, and the leaf corresponding to it is labeled with the corresponding class. The construction of the decision tree strongly depends on how a test attribute X is selected. C4.5 algorithm uses information entropy evaluation function as the selection criteria.

The entropy evaluation function is arrived at through the following steps.

Step 1: Calculate *Info(S)* to identify the class in the training set S.

$$Info(S) = -\sum_{i=1}^{K}\{[freq(C_i, S/|S|)]\log_2[freq(C_i, S/|S|)]\} \tag{6}$$

where, $|S|$ is the number of cases in the training set. C_i is a class, $I = 1,2,3,\ldots.K$ is the number of classes and $freq(C_i,S)$ is the number of cases included in C_i.

Step 2: Calculate the expected information value, $infoX(S)$ for test X to partition samples in S.

$$InfoX(S) = -\sum_{i=1}^{K}[(|S_i|/|S|)Info(S_i)] \tag{7}$$

where K is the number of outputs for test X, S_i is a subset of S corresponding to ith output and is the number of cases of subset S_i.

Step 3: Calculate the information gain

$$Gain(X) = Info(S) - Info_X(S) \tag{8}$$

Step 4: Calculate the partition information value *Splitinfo(X)* acquiring for S, partitioned into L subsets.

$$SplitInfo(X) = -\frac{1}{2}\sum_{i=1}^{L}\left[\frac{|S_i|}{|S|}\log_2\frac{|S_i|}{|S|} + \left(1-\frac{|S_i|}{|S|}\right)\log_2\left(1-\frac{|S_i|}{|S|}\right)\right] \tag{9}$$

Step 5: Calculate the gain ratio

$$GainRatio(X) = Gain(X) - SplitInfo(X) \tag{10}$$

The *GainRatio(X)* compensates for the weak point of *Gain(X)*, which represents the quantity of information provided by X in the training set. Therefore, an attribute with the highest *GainRatio(X)* is taken as the root of the decision tree.

4.2. Application of decision tree for feature selection

A standard tree induced with c4.5 consists of a number of branches, one root, a number of nodes and a number of leaves. One branch is a chain of nodes from root to a leaf; and each node involves one attribute. The occurrence of an attribute in a tree provides the information about the importance of the associated attribute (Peng, Flach, Brazdil, & Soares, 2002). J48 Algorithm (A WEKA implementation of c4.5 algorithm) is a widely used one to construct decision trees (Quinlan, 1996). The procedure of forming the decision tree and exploiting the same for feature selection is explained as follows:

1. The set of features available at hand forms the input to the algorithm; the output is the decision tree.
2. The decision tree has leaf nodes, which represent class labels, and other nodes associated with the classes being classified.
3. The branches of the tree represent each possible value of the feature node from which they originate.
4. The decision tree can be used to classify feature vectors by starting at the root of the tree and moving through it until a leaf node, which provides a classification of the instance, is identified.
5. At each decision node in the decision tree, one can select the most useful feature for classification using appropriate estimation criteria. The criterion used to identify the best feature invokes the concepts of entropy reduction and information gain (Yang et al., 2007a).

The decision tree algorithm has been applied to the problem under discussion. Input to the algorithm is the set of statistical features. The output is the decision tree, a sample decision tree obtained in selecting the statistical features of 12 different faults for 500 rpm is shown in Fig. 5.

It is clear that the top node is the best node for classification. The other features in the nodes of decision trees appear in descending order of importance. It is to be stressed here that only features that contribute to the classification appear in the decision tree and others do not. Features, which have less discriminating capability, can be consciously discarded by deciding on the threshold. This concept is made use for selecting those features which contribute more towards classification. The algorithm identifies the good features for the purpose of classification from the given training data set, and thus reduces the domain knowledge required to select good features for pattern classification problem. Here, the features standard error, median, standard deviation, sample variance and skewness were selected for further study.

5. Support vector machine (SVM)

The next logical step is classification using a classifier. Support vector machines is used as the classifier here. It is a new generation learning system based on statistical learning theory. SVM belongs to the class of supervised learning algorithms in which the learning machine is given a set of features (or inputs) with the associated labels (or output values). Each of these features can be looked upon as a dimension of a hyper-plane. SVMs construct a hyper-plane that separates the hyper-space into two classes (this can be extended to multi-class problems). While doing so, SVM algorithm tries to achieve maximum separation between the classes (see Fig. 6). Separating the classes with a large margin minimizes the expected generalization error. By 'minimum generalization error', we mean that when a new set of features (that is data points with unknown class values) arrive for classification, the chance of making an error in the prediction (of the class to which it belongs) based on the learned classifier (hyper-plane) should be minimum. Intuitively, such a classifier is one, which achieves maximum separation-margin between the classes. The above process of maximizing separation leads to two hyper-planes parallel to the separating plane, on either side of it. These two can have one or more points on them. The planes are known as 'bounding planes' and the distance between them is called as 'margin'. By SVM

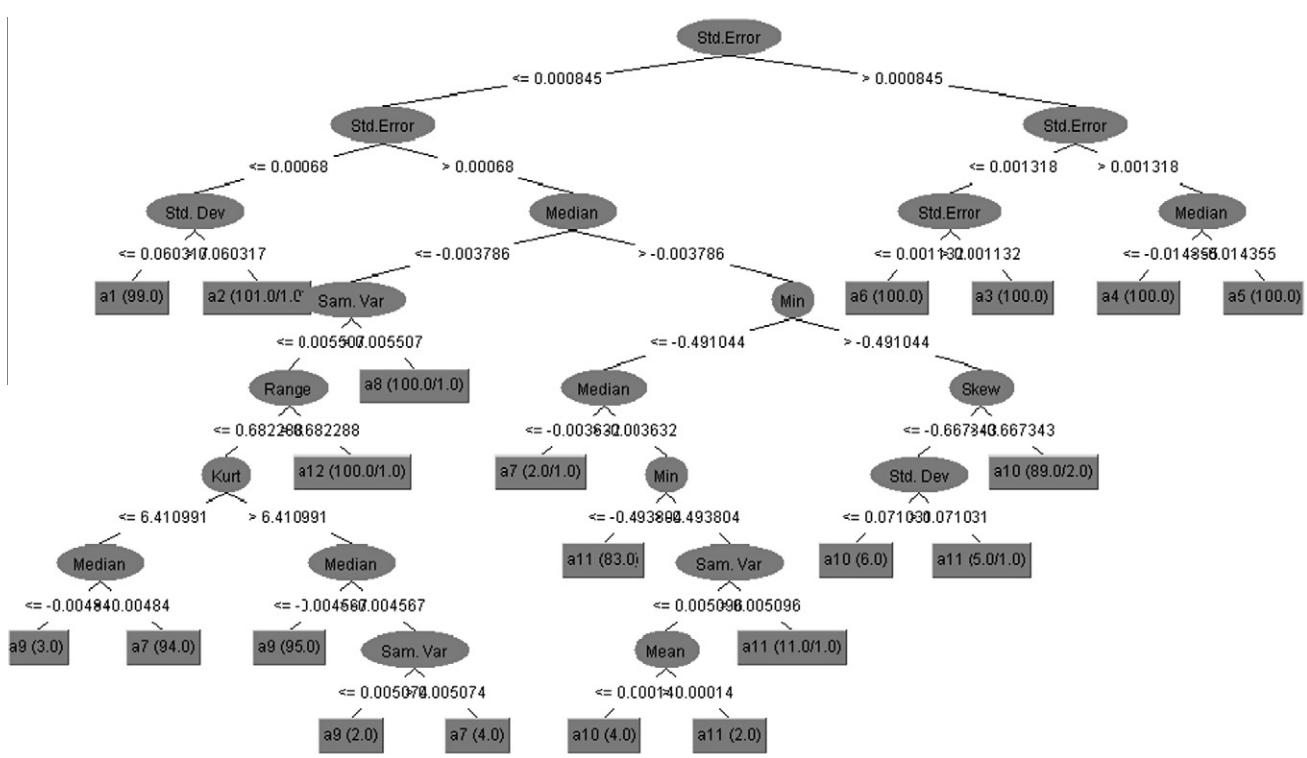

Fig. 5. Decision tree.

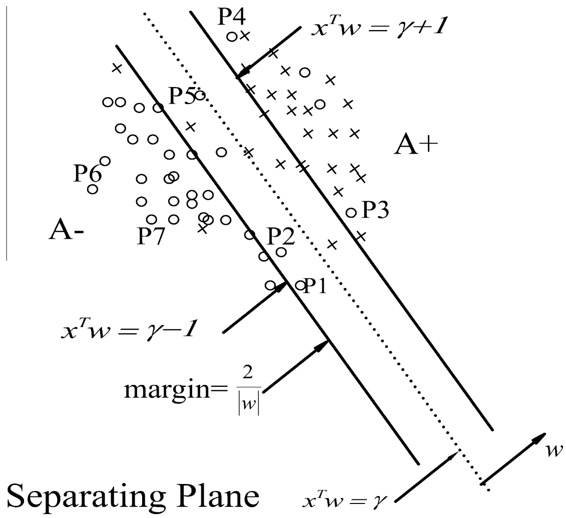

Fig. 6. Standard SVM Classifier.

'learning', we mean, finding a hyper-plane, which maximizes the margin and minimizes the misclassification error. The points lying beyond the bounding planes are called support vectors. The data points P1, P2, P3, P4, and P5 belonging to A− are support vectors, but P6, P7 are not. Same facts hold good for class A+. These points play a crucial role in the theory and hence the name support vector machines. Here, by 'machine', we mean an algorithm. In the formulation, 'A' is a $m \times n$ matrix whose elements belong to real space, 'D' is $m \times 1$ matrix representing class label (+1 and −1), 'e' is a vector of ones and 'v' is a control parameter that defines the weight of error minimization and bounding plane separation in the objective function. 'w' is orientation parameter and 'γ' is location parameter (location relative to origin) of separating hyper plane.

$$\left.\begin{aligned} &\min_{(w,\gamma,y)\in R^{n+1+m}} \quad ve'y + \tfrac{1}{2}w'w \\ &s.t.\ D(Aw - e\gamma) + y \geqslant e \\ &\quad y \geqslant 0 \end{aligned}\right\} \tag{11}$$

where, $A \in R^{m\times n}$, $D \in \{-1,+1\}^{m\times l}$, $e = 1^{m\times l}$.

Vapnik (1999) has shown that if the training features are separated without errors by an optimal hyper-plane, the expected error rate on a test sample is bounded by the ratio of the expectation of the support vectors to the number of training vectors. The smaller the size of the support vector set, more general the above result. Further, the generalization is independent of the dimension of the problem. In case such a hyper-plane is not possible, the next best is to minimize the number of misclassifications whilst maximizing the margin with respect to the correctly classified features.

After training, for any new set of features prediction of its class is possible using the decision function as given below, which is a function of 'w' and 'γ'. It is called testing.

$$f(x) = sign(w^T x - \gamma) \tag{12}$$

If the value of $f(x)$ is positive then new set of features belongs to class A+; otherwise it belongs to class A−.

5.1. c-SVC classification

Given training vectors $x_i \in R^n$, i = 1,...,l, in two classes, and a vector $y_i \in R^l$ such that $y_i \in \{1,-1\}$, c-SVC (Boser, Guyon, & Vapnik, 1992; Cortes & Vapnik, 1995) solves the following primal problem:

$$\begin{aligned} &\min_{w,b,\xi} \quad \frac{1}{2}w^T w + C\sum_{i=1}^{l}\xi_i \\ &\text{subject to } y_i(w^T\phi(x_i) + b) \geqslant 1 - \xi_i \\ &\qquad\qquad \xi \geqslant 0, \quad i = 1,\ldots l \end{aligned}$$

Its dual is

$$\begin{aligned} &\frac{\min}{\alpha} \quad \frac{1}{2}\alpha^T Q\alpha - e^T\alpha \\ &\text{subject to } \ y^T\alpha = 0 \\ &0 \leqslant \alpha_i \leqslant C, \quad i = 1,\ldots,l \end{aligned}$$

where e is the vector of all ones, $C > 0$ is the upper bound, Q is a l by l positive semidefinite matrix, $Q_{ij} = y_i y_j K(x_i x_j)$, and $K(x_i x_j) = \phi(x_i)^T\phi(x_j)$ is the kernel. Here training vectors x_i are mapped into a higher (maybe infinite) dimensional space by the function ϕ.

The decision function is

$$\text{sgn}\left(\sum_{i=1}^{l} y_i\alpha_i K(x_i, x) + b\right) \tag{13}$$

5.2. v-Support vector classification

The v-support vector classification (Schölkopf, Smola, Williamson, & Bartlett, 2000) uses a new parameter v which controls the number of support vectors and training errors. The parameter $v \in (0,1]$ is an upper bound on the fraction of training errors and a lower bound of the fraction of support vectors.

Given training vectors $x_i \in R^n$, $i = 1,\ldots,l$, in two classes and a vector $y \in R^l$ such that $y_i \in \{1,-1\}$, the primal form considered is:

$$\begin{aligned} &\min_{w,b,\xi,\rho} \quad \frac{1}{2}w^T w - v\rho + \frac{1}{l}\sum_{i=1}^{l}\xi_i \\ &\text{subject to } y_i(w^T\phi(x_i) + b) \geqslant \rho - \xi_i \\ &\qquad\qquad \xi_i \geqslant 0, \quad i = 1,\ldots,l, \ \rho \geqslant 0 \end{aligned}$$

The dual is

$$\begin{aligned} &\frac{\min}{\alpha} \quad \frac{1}{2}\alpha^T Q\alpha \\ &\text{Subject to } 0 \leqslant \alpha_i \leqslant 1/l, \quad i = 1,\ldots,l, \ e^T \geqslant v, \ y^T\alpha = 0 \\ &\text{Where} \quad Q_{ij} = y_i y_j K(x_i, x_j) \end{aligned}$$

The decision function is

$$\text{sgn}\left(\sum_{i=1}^{l} y_i\alpha_i K(x_i, x) + b\right) \tag{14}$$

5.3. Application of SVM for feature classification

For each class (shaft and bearing condition), features consisting of 200 feature value sets were collected from the experiment. The best features were selected using decision tree. One hundred samples in each class were used for training the SVM with selected features and 100 samples were reserved for testing. The results are discussed in Section 6.

6. Results and discussion

Classification is a two phase process: training and testing. Training is the process of learning to label from the examples. Training can be supervised mode or unsupervised mode. Here, supervised mode is used for training. Testing is the process of checking how well the classifier has learnt to label the unseen examples. The four

Table 1
Performance of SVM kernel function.

Sl. no	SVM kernel function	Classification efficiency (%)							
		Speed 500 rpm		Speed 700 rpm		Speed 900 rpm		Speed 1100 rpm	
		c-SVC	nu-SVC	c-SVC	nu-SVC	c-SVC	nu-SVC	c-SVC	nu-SVC
1	Linear	98.833	98.000	98.083	95.583	99.167	98.917	99.833	99.667
2	Three degree polynomial	98.833	98.917	98.000	93.917	99.500	97.833	99.833	99.833
3	Radial basis function (RBF)	99.333	99.333	98.250	97.833	99.417	99.333	99.917	99.833
4	Sigmoid	98.917	98.250	98.083	96.000	99.167	94.917	99.833	98.500

Table 2
Confusion matrix of SVM using RBF of c-SVC model for 500 rpm.

	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12
a1	100	0	0	0	0	0	0	0	0	0	0	0
a2	0	100	0	0	0	0	0	0	0	0	0	0
a3	0	0	100	0	0	0	0	0	0	0	0	0
a4	0	0	0	100	0	0	0	0	0	0	0	0
a5	0	0	0	0	100	0	0	0	0	0	0	0
a6	0	0	0	0	0	100	0	0	0	0	0	0
a7	0	0	0	0	0	0	96	0	1	3	0	0
a8	0	0	0	0	0	0	0	100	0	0	0	0
a9	0	0	0	0	0	0	1	0	99	0	0	0
a10	0	0	0	0	0	0	1	0	0	98	1	0
a11	0	0	0	0	0	0	0	0	0	0	100	0
a12	0	0	0	0	0	0	1	0	0	0	0	99

different kernel functions, such as linear function, three degree polynomial function, RBF and sigmoid function of c-SVC and nu-SVC models of SVM were used for classification. The classification efficiency for the four different functions of two SVM models is given in Table 1.

From Table 1, both the SVM models yielded classification accuracy in excess of 93.9% of all the kernel function tested; RBF provided accuracy in excess of 97.833% in fault classification with both c-SVC and nu-SVC. When comparing the RBF classification efficiency, c-SVC model is better than nu-SVM model for fault diagnosis of rotational mechanical system. The sample testing result of the classifier for 500 rpm is presented in the form of confusion matrix in Table 2.

The interpretation of the confusion matrix is as follows:

- The diagonal elements in the confusion matrix show the number of correctly classified instances. In the first row, the first element shows number of data points belongs to 'good' class (a1) and classified by classifier as 'good'.
- The second element shows the number of data points belongs to 'good' class but misclassified as unbalanced shaft good bearing (a2) and the third element shows the number of data points misclassified as good shaft IRF bearing (a3) and so on.

From the confusion matrix (Table 2), one can note that only the data points of a7 (shaft bent good bearing), a9 (shaft bent with ORF bearing), a10 (shaft bent with unbalancing mass good bearing) and a12 (shaft bent with unbalancing mass and ORF bearing) are misclassified. SVM finds it difficult to discriminate between a7 and a10. Misclassification of 4% in a7 and 2% in a10 brings down the diagnostic ability of the SVM; however, the overall classification efficiency (99.333%) is reasonably good.

7. Conclusion

Fault diagnosis of shaft and bearings is one of the core research areas in the field of condition monitoring of rotating machines. Many researchers reported the fault diagnosis of either shaft or bearing, but here both the shaft and the bearing have been considered. The statistical features for 12 different conditions were extracted from the vibration signals. Decision tree is used to select the best features. The best features were classified using four different SVM kernel functions of two SVM model in support vector machine. The RBF of c-SVC model gives the better classification efficiency for four different speeds. From the above result one can conclude that c-SVC model of SVM classifier with RBF kernel function is a good candidate for fault diagnosis of rotational mechanical system.

References

Abbasion, S., Rafsanjani, A., Farshidianfar, A., & Irani, N. (2007). Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. *Mechanical Systems and Signal Processing, 21*(7), 2933–2945.

Boser, B. E., Guyon, I., & Vapnik, V. (1992). A training algorithm for optimal margin Classifers. In Proceedings of the fifth annual workshop on computational learning theory (pp. 144–152).

Breiman, L., Friedman, J., Olshen, R., et al. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth International Group.

Burgess, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery, 2*, 955–974.

Cempel, C. (1991). Condition evolution of machinery and its assessment from passive diagnostic experiment. *Mechanical System and Signal Processing, 5*(4), 317–326.

Cortes, C., & Vapnik, V. (1995). Support-vector network. *Machine Learning, 20*, 273–297.

Jack, L. B., & Nandi, A. K. (2000). Comparison of neural networks and support vector machines in condition monitoring application. In Proceedings of COMADEM 2000, Houston, TX, USA, 2000 (pp. 721–730).

Kalkat, M., Yildirim, S., & Uzmay, I. (2005). A Design of artificial neural networks for rotor dynamics analysis of rotating machine systems. *Mechatronics*(15), 573–588.

Peng, Y. H., Flach, P. A., Brazdil, P., & Soares, C. (2002). Decision Tree-based data characterization for meta-learning. In ECML/PKDD-2002 workshop IDDM-2002. Finland: Helsinki.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. San Mateo, CA: Morgan, Kaufmann.

Quinlan, J. R. (1996). Improved use of continuous attributes in C4.5. *Journal of Artificial Research*(4), 77–90.

Salido, M. F., & Shuta, M. (2004). A comparison of two learning mechanisms for the automatic design of fuzzy diagnosis systems for rotating machinery. *Applied Soft Computing, 4*, 413–422.

Samanta, B., & Al-Baulshi, K. R. (2003). Artificial neural network based fault diagnostics of rolling element bearings using time domain features. *Mechanical Systems and Signal Processing, 17*(2), 317–328.

Schölkopf, B., Smola, A., Williamson, R. C., & Bartlett, P. L. (2000). New support vector algorithms. *Neural Computation, 12*, 1207–1245.

Shi, X. Z., Xu, Z. Q., & Xu (1988). A Study on the automatic recognition of vibration signal for ball bearing faults – The FFTAR feature extraction and classification methods. In *Proceedings of ieee international workshop on applied time series analysis* (pp. 318–321). World Scientific.

Smalley, A. J., Baldwin, R. M., Mauney, D. A., & Millwater, H. R. (1996). Towards risk based criteria for rotor vibration. In International proceedings of the institute of mechanical engineers—Vibrations in rotating machinery (pp. 517–527).

Sugumaran, V., Muralidharan, V., & Ramachandran, K. I. (2007). Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing. *Mechanical Systems and Signal Processing, 21*(2), 930–942.

Sugumaran, V., Sabareesh, G. R., & Ramachandran, K. I. (2008). Fault diagnostics of roller bearing using kernel based neighborhood score multi-class support vector machine. *Expert Systems with Applications, 34*(4), 3090–3098.

Su, YT., & Lin, SJ. (1992). On initial fault-detection of a tapered roller bearing—Frequency-domain analysis. *Journal of Sound and Vibrations, 155*(1), 75–84.

Sun, W., Chen, J., & Li, J. (2007). Decision tree and PCA-based fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing, 1*(21), 1300–1317.

Vapnik, V. N. (1999). *The nature of statistical learning theory*. Springer-Verlag. pp. 138–146.

Winder, R. L., & Littmann, W. E. (1976). *Bearing damage analysis*. National Bureau of Standard Publication.

Yang, Y., Yu, D., & Cheng, J. (2007a). A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM. *Measurement, 40*(9–10), 943–950.

Yang, J., Zhang, Y., & Zhu, Y. (2007b). Intelligent fault diagnosis of rolling element bearing based on SVMs and fractal dimension. *Mechanical Systems and Signal Processing, 21*(5), 2012–2024.