Journal of
Mechanical
Science and
Technology

Journal of Mechanical Science and Technology 21 (2007) 607~615

A Comparison Study of Support Vector Machines and Hidden Markov Models in Machinery Condition Monitoring

Qiang Miao, Hong-Zhong Huang*, Xianfeng Fan

School of Mechatronics Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, 610054, P. R. China

(Manuscript Received August 31, 2006; Revised March 16, 2007; Accepted March 16, 2007)

Abstract

Condition classification is an important step in machinery fault detection, which is a problem of pattern recognition. Currently, there are a lot of techniques in this area and the purpose of this paper is to investigate two popular recognition techniques, namely hidden Markov model and support vector machine. At the beginning, we briefly introduced the procedure of feature extraction and the theoretical background of this paper. The comparison experiment was conducted for gearbox fault detection and the analysis results from this work showed that support vector machine has better classification performance in this area.

Keywords: Pattern recognition; Feature extraction; Hidden markov model; Support vector machine

1. Introduction

Condition monitoring and classification of machinery is a rapidly developing area of machine learning and pattern recognition, especially with the application of artificial intelligence techniques such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM). In the procedure of condition monitoring and classification, vibration data are commonly chosen because of their properties of easy collection, abundant information, and relatively low overhead. However, raw vibration data cannot be directly used as input of classification system, since these signals are usually contaminated by noises. Signal processing and feature extraction are necessary steps that can eliminate noises and enhance useful information related to machine conditions. The basic idea of signal processing and feature extraction is to transform the original data from one domain (usually in time domain) to another space (usually in time-frequency domain) with some mathematical tools. Through this transform, important information can be intensified and identified.

In recent years, a lot of feature extraction methods have been proposed in machinery condition monitoring problem utilizing vibration data. For example, Xu and Ge (2004) proposed an intelligent fault diagnosis system in monitoring stamping process using wavelet packet, in which features are acquired from time marginal energy and frequency marginal energy. Tahk and Shin (2002) proposed a diagnosis algorithm to detect the defective rollers based on the frequency analysis of web tension signals, which is to use the characteristic features (root mean square, peak value, power, spectral density) of tension signals for the identification of the faulty rollers and the diagnosis of the degree of fault in the rollers. In addition, the selection of time-domain signal statistics, such as mean, root mean square, variance, skewness, kurtosis and normalized higher order central moments, is another way of feature extraction (Samanta, 2004) to

*Corresponding author. Tel.: +86 28 83206916, Fax.: +86 28 83206828
E-mail address: hzhuang@uestc.edu.cn

distinguish between normal and defective conditions of pump. In this paper, we realized this step through the extraction of Lipschitz exponent function proposed in (Miao, 2005). During next section, a brief introduction of our feature extraction method will be given.

The research presented in this paper is to compare the classification performance of Hidden Markov Models (HMM) and SVM in gear fault detection. HMM is based on statistical learning theory and is predominantly applied in current speech recognition systems (Huo et al., 1995) because it can normalize the time-variation of the speech signal and characterize the speech signal statistically and optimally. Moreover, application of HMM in condition monitoring is another popular research topic including stamping process monitoring (Xu and Ge, 2004), tool wear monitoring (Ertunc et al., 2001), and bearing fault detection (Ocak and Loparo, 2001), etc. Support Vector Machine, which is a relatively new technique based on statistical learning theory, is gaining interests of researchers in the areas of machine learning, computer vision and pattern recognition (Burges, 1998; Yuan, 2006). The main difference between HMM and SVM is in the principle of risk minimization (RM) (Vapnik, 1999; Samanta, 2004). In case of HMM, empirical risk minimization (ERM) is used, which is the simplest induction principle whereas in SVM, structural risk minimization (SRM) is used as induction principle. The difference in RM leads to better generalization performance for SVM than HMM.

In this paper, we investigate the classification performance of these two methods in machinery condition monitoring. An example of gear fault detection is given to compare their performance from the aspect of classification accuracy. The rest of the paper is organized as follows. Section 2 introduces the procedure of wavelet transform and extraction of Lipschitz exponent function. Theoretical background of SVM and HMM are described in Section 3. In Section 4, a comparison study of SVM and HMM based classifiers is conducted in solving gearbox condition monitoring problem. Conclusion from this research is drawn in Section 5.

2. Feature extraction in condition monitoring

2.1 Diagram of proposed condition monitoring system

Generally, a condition monitoring system of machinery can be divided into three subsystems: signal processing unit, feature extraction unit and condition classification unit. In this paper, vibration data recorded from a gearbox case is used as the example for the comparison study. Raw vibration data is processed with wavelet transform and the output is a time-frequency representation of signal. In feature extraction unit, the Lipschitz exponent function is extracted from time-frequency representation. Finally, a comparison between SVM based classification system and HMM based one is conducted. Figure 1 is a flow chart of condition monitoring procedure in this research.

2.2 Extraction of lipschitz exponent function with wavelet

Wavelets have been gaining popularity as a choice of multi-scale transform since their first appearance in the early 1990s. Wavelets are mathematical functions that decompose a signal into its constituent parts using a set of wavelet basis functions. The family of basis functions used for wavelet analysis is created by both dilations (scaling) and translations (in time domain) of a mother wavelet, thereby providing both time and frequency information about the signal being analyzed. The resulting coefficients from the wavelet transform of a time domain signal, such as the acceleration response of a structure, can be represented in a two-dimensional time-scale map.

Signals recorded from rotating machinery usually include stationary components coming from periodic

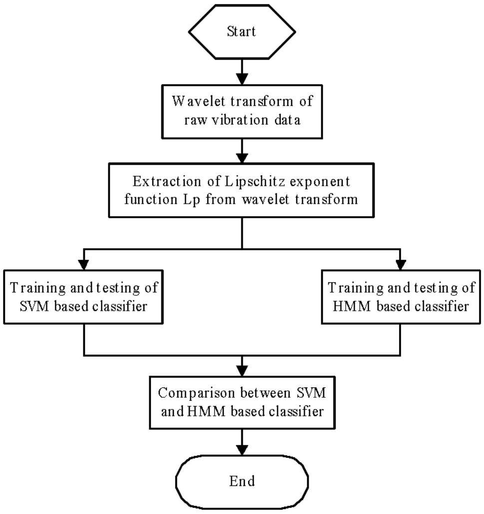

Fig. 1. The diagram of condition monitoring system.

motion of system such as gear meshing. However, the occurrence of machine failure may result in emergence of non-stationary transients masked by noise and normal meshing components of rotating object. A signal processing method that ignores the regular part of a signal and focuses on the transient part has more potential in extracting diagnostic information. In this research, these transient phenomena can be treated as singularities in signal and a local method with good signal to noise characteristics should be applied because of their low energy content. Such a method is the singularity analysis based on wavelet pioneered by Mallat and Hwang (1992). According to this method, most of fault related information in a signal can be captured in the local maxima of the wavelet transform modulus and characterized by Lipschitz exponent, which is a quantitative description of signal's local regularity.

In order to understand the concept of Lipschitz exponent and some new methods derived from it, a strict definition is given here. Assume $f(x)$ to be a finite energy function, that is, $f(x)\in L^2(R)$. The definition of Lipschitz exponent is: $f(x)$ is said to be Lipschitz α at x_0, if and only if there exists two constants A and $h_0 > 0$ and a polynomial $P_n(x)$ of order n, such that

$$f(x) - P_n(x-x_0) \,|\leq A|\, x-x_0|^{\alpha} \text{ for } |x-x_0| < h_0. \quad (1)$$

We call Lipschitz regularity of $f(x)$ and x_0, the superior bound of all values α such that $f(x)$ is Lipschitz α at x_0. In case $f(x)$ is continuously differentiable at a point, it has Lipschitz 1 at this point. If the Lipschitz regularity α of $f(x)$ at $x = x_0$, satisfies $n< \alpha <n+1$, then $f(x)$ is n times differentiable at x_0 but its nth derivative is singular at x_0 and α characterizes this singularity. Here n is a positive integer.

To measure the Lipschitz α, a classical method is to look at the asymptotic decay of the amplitude of function $f(x)$'s Fourier transform. However, the asymptotic decay of a signal's frequency spectrum relates directly to the uniform Lipschitz regularity. It only provides a measure of the minimum global regularity of the function and cannot localize the information along the temporal variable x (Loutridis and Trochidis, 2004), which means it is helpless in handling non-stationary transient signals. On the other hand, if the wavelet $\psi(x)$ has a compact support, the value of wavelet coefficient $Wf(s, x)$ depends on the value of $f(x)$ in a neighborhood, of size proportional to the scale s. The wavelet method for estimating the Lipschitz exponent is similar to that of the Fourier transform. By examining the decay of time-scale map at specific points in time across all scales (frequencies), the local Lipschitz regularity of the signal can be determined.

In order to measure Lipschitz exponent α, we must impose that the wavelet $\psi(x)$ has enough vanishing moments. For example, a wavelet $\psi(x)$ is said to have n+1 vanishing moments, if and only if for all positive integers $k < n+1$, following condition can be satisfied,

$$\int_{-\infty}^{+\infty} x^k \psi(x)dx = 0, for 0 < k < n+1 \quad (2)$$

Eq. (2) guarantees that the wavelet with n+1 vanishing moments is orthogonal to the polynomials of order up to n. If the function $f(x)$ is Lipschitz α at x_0, $n< \alpha <n+1$, then there exists a constant A such that for all points x in a neighbourhood of x_0 and any scales (Mallat and Hwang, 1992),

$$|Wf(s,x)| \leq A(s^{\alpha} + |x - x_0|^{\alpha}) \quad (3)$$

The local Lipschitz α of $f(x)$ at x_0 depends on the decay of $|Wf(s, x)|$ at fine scales in the neighborhood of x_0. The decay can be measured through the local modulus maxima. Define modulus maxima as any point (s_0, x_0) such that $|Wf(s_0, x)|$ is a local maxima at $x = x_0$. If there exists a scale $s_0 > 0$, and a constant C, such that for $x\in (a, b)$ and $s< s_0$, all the modulus maxima of $Wf(s, x)$ belong to a cone defined by (Miao and Makis, 2007)

$$|x - x_0| \leq Cs, \quad (4)$$

Here a and b are boundaries that define a small interval of x. Then at each modulus maximum (s, x) in the cone defined by (4),

$$|Wf(s,x)| \leq Bs^{\alpha}, \quad (5)$$

which is equivalent to

$$\log_2 |Wf(s,x)| \leq \log_2 B + \alpha \log_2 s. \quad (6)$$

Where $B = A(1+C^{\alpha})$.

Using the above mentioned method, singularities in a signal can be detected by examining the evolution of the modulus maxima of the wavelet transform along maxima line. Modulus maxima line is defined by connecting the adjacent modulus maxima points. An alternative to the extraction of the maxima line is to look at the decay of the wavelet modulus across the scales at each time point. Points where large changes occur in the signal (such as singularities) will have large coefficients at certain scales, thus having significant decay. The measure of this decay is the Lipschitz exponent of the signal at a given time point. By examining the change of this exponent in time, singularities can be identified.

Based on the analysis above, Lipschitz exponent function $Lp(x)$ can be derived as follows. The resulting coefficients from wavelet transform of vibration signal generate a two-dimensional time-scale representation which takes a form of matrix. One dimension of the time-scale matrix represents a different scale (s), while the other one (x) denotes a different time point in the signal. Take each column, which represents the frequency spectrum of the signal at the corresponding time point x, and use Linear Regression method to approximately estimate α (Robertson et al., 2003). Finally the Lipschitz exponent function $Lp(x)$ can be achieved which is a function that describes the change of Lipschitz value along x. The choice of Linear Regression method is a simplification of (6) and similar technique is validated by Robertson *et al.* (2003) in structural health monitoring.

3. Theoretical background of SVM and HMM

3.1 Condition classification based on SVM

Support vector machine (SVM) (Vapnik, 1999) is a classifier that estimates decision surfaces directly rather than modeling a probability distribution across the training data. SVMs have demonstrated good performance on several classic pattern recognition problems including machinery condition monitoring (Burges, 1998; Yang et al., 2005; Yuan and Chu, 2006). Figure 2 is an example of the classification of a series of points for two different classes (represented by triangles and squares, respectively). It is a typical 2-class (positive and negative classes) SVM problem in which the points are perfectly separable using a linear decision region. H is a separating hyperplane. H1 and H2 define two hyperplanes and they are parallel to H. The distance between H1 and H2 is called the margin. The closest positive class and negative class points lying on these two hyperplanes are called the support vectors.

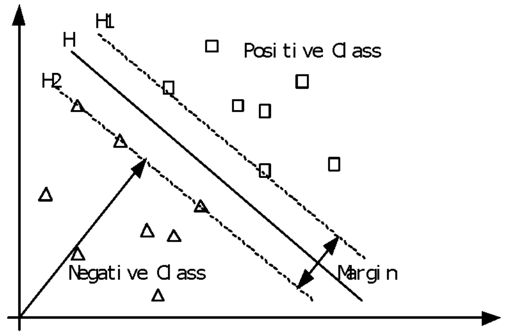

Fig. 2. An example of 2-class classifier by SVM.

In this section, we briefly describe the SVM theory and more details can be found in (Burges, 2004). Assume we have a data set:

$$D=\{(X_i, y_i)\}\ (i=1,\dots l),\ y_i \in \{-1, +1\} \tag{7}$$

where X_i is input sample and y_i is output class. The goal of SVM is to define a hyperplane which divides D, such that all the points in the same class are on the same side of the hyperplane H while maximizing the distance between the two classes and the separating hyperplane. The optimal separating plane is a linear classifier that can be defined as the following decision function:

$$f(x) = \mathrm{sgn}\left\{\sum_{i=1}^{l} \lambda_i y_i X^T X_i + b\right\} \quad (i = 1, \dots, l). \tag{8}$$

Here sgn(*) is the sign function and the Lagarange coefficient λ_i is the solution of the following quadratic programming problem (Yuan and Chu, 2006):

$$\text{Maximize } W(\lambda) = -\sum_{i=1}^{l} \lambda_i + \frac{1}{2}\sum_{i,j}^{l} \lambda_i \lambda_j y_i y_j X_i X_j \tag{9}$$

$$\text{Subject to } \sum_{i=1}^{l} \lambda_i y_i = 0,\ \lambda_i \geq 0$$

where $\lambda = \{\lambda_i\}$. In real-world application, problems typically involve data which can only be separated using a nonlinear decision surface. Optimization on the input data in this case involves the use of a kernel-based transformation. By projecting the original sample space into a high-dimensional eigenspace with a kernel function $K(X, X_i)$, the nonlinear problem becomes linearly separable and the SVM classifier has the classifying function as follows:

$$f(X) = \text{sgn}\left\{\sum_{i=1}^{l} \lambda_i y_i K(X, X_i) + b\right\} \quad (10)$$

Vapnik (1999) provides us several kernel functions like linear, polynomial, Gaussian and Laplacian radial basis functions (RBF). In this research, the Gaussian RBF kernel, given by Eq. (11) is used.

$$K(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2) \quad (11)$$

where σ denotes the width of the Gaussian RBF kernel parameter and can be determined by an iterative process to select an optimum value based on the full feature set.

3.2 Condition classification based on HMM

In the last decade, HMM has attracted the attention of many researchers in pattern recognition such as handwriting, speech and signature verification. The power of an HMM lies in its ability to model the temporal evolution of a signal via an underlying Markov process. Widespread use of HMMs can be attributed to the availability of efficient parameter estimation procedures that involve maximizing the likelihood of the data given the model. This is an iterative method called expectation maximization (EM) algorithm. The EM algorithm provides an iterative framework for maximum likelihood estimation with good convergence property, though it does not guarantee finding the global optimality.

HMM has advantages that allow it to provide solutions by modeling and learning by itself, even if it does not have exact knowledge about the problem to be solved. HMMs are divided into continuous and discrete models according to their probability density distribution. In this paper, discrete HMMs are selected because they can represent any distribution as no assumptions are made regarding the underlying distribution.

HMM is represented by a graph structure that consists of N nodes, called hidden states, and arcs that represent transitions between nodes. Figure 3 is an example of HMM model with N=3 hidden states. The shadow nodes in the figure represent observations from corresponding hidden states. The observation symbol probability distribution models spatial characteristics, and the state transition probability distribution models time evolution characteristics. HMM states are not directly observable, but can be inferred through a sequence of observed symbols. To describe the HMM formally, the following model notations will be used.

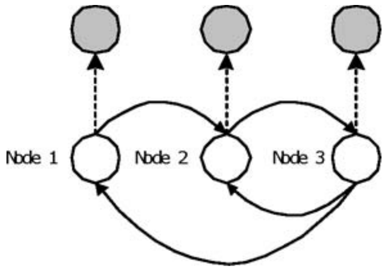

Fig. 3. A simple example of HMM with 3 hidden states.

- set of hidden states: $S = \{S_1, S_2, ..., S_N\}$, where N is the number of states in HMM,
- state transition probability distribution: $A = \{a_{ij}\}$, where $a_{ij} = P[q_{t+1} = S_j \mid q_t = S_i]$, for $1 \le i, j \le N$,
- set of observation symbols: $V = \{v_1, v_2, ..., v_M\}$, where M is the number of observation symbols per state,
- observation symbol probability distribution: $B = \{b_j(k)\}$, where $b_j(k) = P[v_k \text{ at } t \mid q_t = S_j]$, for $1 \le j \le N$, $1 \le k \le M$.
- initial state probability distribution: $\pi = \{\pi_i\}$, where $\pi_i = P[q_1 = S_i]$, for $1 \le i \le N$.

Where q_t represents the hidden state at time t. An HMM can be represented by the compact notation $\lambda = (A, B, \pi)$. HMM modeling involves choosing the number of hidden states, N, the number of discrete symbols, M, and the specification of three probability distributions A, B, and π.

4. Comparison Study between SVM and HMM

4.1 Experimental set-up

Vibration data used in this paper are measured from a gearbox driven by an electrical motor under laboratory environment. This is a single stage gear reduction unit mounted on a mechanical diagnostic test bed (MDTB, see Fig. 4). The MDTB is functionally a motor-drive train-generator test stand (Byington and Kozlowski, 2000). The gearbox is driven at a set input speed using a 30 Hp, 1750 rpm AC (drive) motor, and the torque is applied by a 75 Hp, 1750 rpm AC (absorption) motor. The maximum speed and torque are 3500 rpm and 225 ft-lbs respectively. In this test run, the shaft speed is kept constant at 1750rpm. The variation of the torque is accomplished by a vector unit capable of controlling the current output of the absorption motor. The MDTB is highly efficient because the electrical power that is generated by the absorber is fed back to the driver motor. The mechanical and electrical losses are sustained by a small fraction of wall power. The MDTB has the capability of testing single and double reduction industrial gearboxes with ratios from about 1.2:1 to 6:1. The general information of gearbox is shown in Table 1.

The experiment started with a brand new gearbox under 100% of rated workload. After sometime, it increased to 200% of rated workload and ran until some criteria of gearbox health condition were satisfied. However, the gear failure occured before the testrig was stopped. During this procedure, vibration data were recorded occasionally with a period of 10 seconds at a sampling rate of 20 kHz. Each time we collected vibration data (10 seconds), it was saved as a data file and numbered consequently. Therefore, there are 148 data files, among which 12 of them (file 1 to 12) are under 100% rated workload and normal conditon, 25 of them (file 13 to 37) under 200% rated workload and normal conditon, and the remaining 111 of them (file 38 to 148) under 200% rated workload and failure condition. Figure 5 is a graphic demonstration of data history in this experiment.

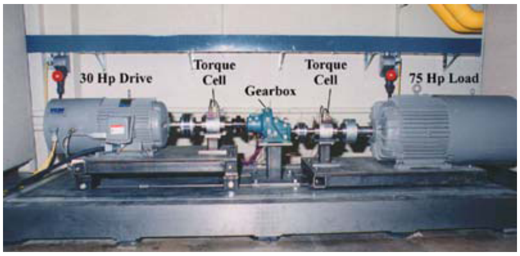

Fig. 4. Mechanical diagnostic test bed.

Table 1. Gearbox information in this experiment.

Gearbox ID#	DS3S0150XX
Make	Dodge APG
Model	R86001
Rated Input Speed	1750 rpm
Gear Ratio	1.533
Contact Ratio	2.388
Number of Teeth (Driven gear)	46
Number of Teeth (Pinion gear)	30

Each data file has a length of 10 seconds (200,000 sampling points), and one revolution of gear contains 1052 sampling points, which is calculated based on mechanical specifications of the gearbox. Thus, each data file contains around 190 periods of revolution and can be divided into 10 bins (10×19=190), which means we can extract 10 *Lp* functions from each data files. These *Lp* functions will be used as input features of classifiers.

As we introduced previously, there are 148 gearbox vibration data files, which can be divided into three sets of data. That is, 12 of them are under 100% rated workload and normal condition (Dataset A), 25 of them under 200% rated workload and normal condition (Dataset B), 111 of them under 200% workload and failure condition (Dataset C). Since Dataset B and C are under same working environment and different machine conditions, we select data from B and C for the training and testing of corresponding classifiers without consideration of Dataset A. We choose 25 data files from Dataset B and 30 from Dataset C to form normal (N1) and failure (F1) datasets, respectively. In addition, we select another 30 data files from Dataset C and name it as F2. Therefore, there are three datasets, N1, F1 and F2. The difference between F1 and F2 is that F1 is selected from data files that the gearbox is under early failure stage (file number between 38 and 92) and F2 is selected from data files that the gearbox is under heavy failure stage (file number between 93 and 148). Figure 6 is a graph to show the visual difference of Lipschitz exponent functions from these three datasets.

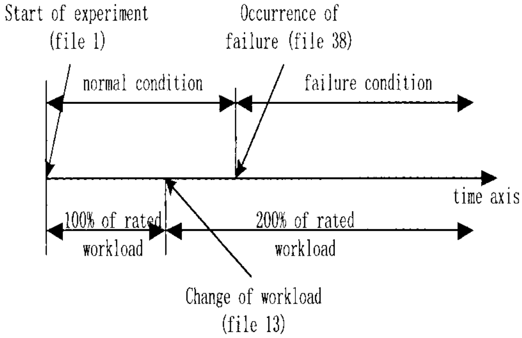

Fig. 5. Description of vibration data history.

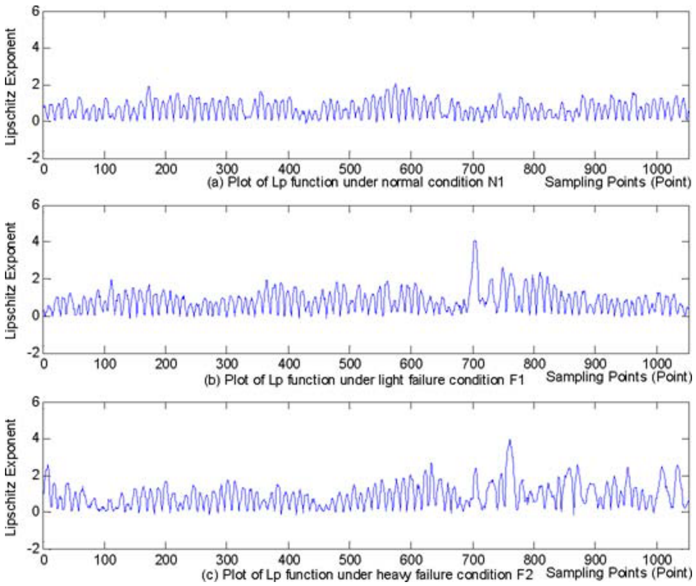

Fig. 6. Plot of Lipschitz exponent functions in three datasets (N1, F1, F2).

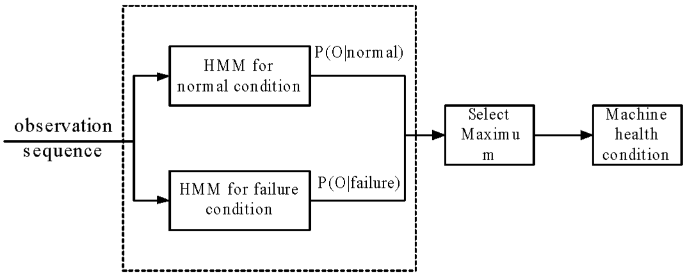

Fig. 7. Structure of HMM based classification system.

4.2 Specification of SVM and HMM based classifiers in this comparison

In SVM modeling, since this is a two-class problem, one SVM model is enough for classification. In this paper, we used SVM Matlab toolbox developed by Rakotomamonjy and Canu (2005). The Gaussian RBF is chosen as the kernel function and the parameter σ is 0.5, which is based on n-fold cross-validation procedure. In this procedure, 4 data files from N1 and F1 respectively are split into 4 partitions and σ is selected from 0.1 to 0.9 with step size 0.1. Although the convergence of RBF kernels is typically slower than that of polynomial kernels, RBF kernels often deliver better performance and they are extremely popular (Shao and Chang, 2005). In the training step, we train the SVM based classifier with Lipschitz exponent functions extracted from 4 data files of N1 and F1 respectively and the remaining are used for testing.

In HMM modeling, two HMM models are established, one for normal condition and the other for failure condition. In this case, the hidden states of each HMM do not have certain physical meanings. In testing procedure, compute the log-likelihood for each model given the input observation (features) and make decision based on the model that has larger log-likelihood. The number of hidden states in each model is 5, which is based on the evaluation of HMM-based system performance with different hidden states. Figure 7 shows the structure of HMM based classification system.

4.3 Comparison results

This is a typical two-class condition classification problem because we only consider two machinery conditions, namely normal and failure. In vibration

Table 2. Performance comparison of HMM and SVM based classifiers.

Dataset	HMM based classifier		SVM based classifier	
	Test success (%)	Training time (s)	Test success (%)	Training time (s)
N1	85.71	12.371	85.71	4.783
F1	88.46	13.103	92.31	5.892
F2	76.67		83.33	

signal processing, comblet and wavelet transform has been applied and details can be found in (Miao, 2005). Since the signal in each data file is partitioned into 10 bins, wavelet transform is applied for these 10 pieces of signals. In feature extraction, 10 Lipschitz exponent functions *Lp* corresponding to the wavelet transform of 10 pieces of signals are extracted from one data file. Therefore, there are 250, 300, 300 feature vectors in dataset N1, F1 and F2, respectively. In the training step, we use 40 feature vectors from 4 data files in N1 and F1 datasets. In the testing step, all the remaining data are used for validation.

The performance of the classifiers is evaluated using the detection rate, or test success percentage. Table 2 is the analysis results from this study. Theoretically, the performance of SVM based classifier should be better than HMM based classifier since the former one is based on the principle of structural risk minimization (Justino et al., 2005). The comparison shows that SVM has better performance than HMM in the testing dataset F1 (92.31% with SVM based classifier v.s. 88.46% with HMM based classifier) although the difference is not very distinct (since they have same detection rate in N1 dataset, 85.71%). In addition, when we use Dataset F2 to test the classifiers, the drop of performance of HMM based classifier (from 88.46% to 76.67%) is greater than that of SVM (from 92.31% to 83.33%). This is an interesting phenomenon and it demonstrates that SVM has better generalization than HMM because with the aggravation of fault, the previously trained classifier may not be accurate enough. From this aspect, another problem is proposed for the adaptive training of classifier. Furthermore, the computation (for a PC with Pentium IV processor of 1.7GHz and 512Mb RAM) for training the classifiers are also shown, and SVM has better efficiency. However, it should be mentioned that the difference in computation time should not be very important if the training is done off-line.

5. Conclusions and future work

The paper presents a brief comparison of HMM and SVM in machinery condition monitoring. This is an investigation of pattern recognition techniques for the design of an efficient and robust classification system. Until now, based on the analysis results we can see that SVM has better recognition performance and generalization. In the future research, we will investigate the design of adaptive classifier because it is a realistic problem in machinery condition monitoring due to the deterioration of machine condition.

Acknowledgments

This research was partially supported by the National Natural Science Foundation of China under contract number 60672165. We are most grateful to the Applied Research Laboratory at the Pennsylvania State University and the Department of the Navy, Office of the Chief of Naval Research (ONR) for providing the vibration data used to develop this work. We also appreciate very much the valuable and constructive comments from the anonymous referees who greatly helped to improve this work.

References

Burges, C. J. C., 1998, "A tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, Vol. 2, pp. 121~167.

Byington, C. S., Kozlowski, J. D., 2000, "Transitional Data for Estimation of Gearbox Remaining Useful Life," *Mechanical Diagnostic Test Bed Data, Applied Research Laboratory*, The Pennsylvania State University.

Ertunc, H. M., Loparo, K. A., Ocak, H., 2001, "Tool Wear Condition Monitoring in Drilling Operations Using Hidden Markov Models(HMM)," *International Journal of Machine Tools & Manufacture*, Vol. 41, pp. 1363~1384.

Huo, Q., Chan, C., Lee, C. H., 1995, "Baysian Adaptive Learning of the Parameters of Hidden Markov Model for Speech Recognition," *IEEE Transaction on Speech and Audio Processing*, Vol. 3, No. 5, pp. 334~345.

Justino, E. J. R., Bortolozzi, F., Sabourin, R., 2005, "A Comparison of SVM and HMM Classifiers in the Off-Line Signature Verification," *Pattern Recognition Letters*, Vol.26, pp. 1377~1385.

Loutriis, S., Trochidis, A., 2004, "Classification of Gear Faults Using Hoelder Exponents," *Mechanical Systems and Signal Processing*, Vol. 18, pp. 1009~1030.

Mallat, S., Hwang, W. L., 1992, "Singularity Detection and Processing with Wavelets," *IEEE Transactions on Information Theory*, Vol. 38, No. 2, pp. 617~643.

Miao, Q., Makis, V., 2007, "Condition Monitoring and Classification of Rotating Machinery Using Wavelets and Hidden Markov Models," *Mechanical Systems and Signal Processing*, Vol. 21, No. 2, pp. 840~855.

Miao, Q., 2005, "Application of Wavelets and Hidden Markov Model in Condition-Based Maintenance," *Ph.D Thesis*, University of Toronto.

Ocak, H., Loparo, K. A., 2001, "A New Bearing Fault Detection and Diagnosis Scheme Based on Hidden Markov Modeling of Vibration Signals," *IEEE ICASSP 2001*, Vol. 5, pp. 3141~3144.

Rakotomamonjy, A., Canu, S., 2005, http://asi.insa-rouen.fr/~arakotom/toolbox/index.html.

Robertson, A. N., Farrar, C. R., Sohn, H., 2003, "Singularity Detection for Structural Health Monitoring Using Holder Exponents," *Mechanical Systems and Signal Processing*, Vol. 17, No. 6, pp. 1163~1184.

Samanta, B., 2004, "Gear Fault Detection Using Artificial Neural Networks and Support Vector Machines with Genetic Algorithms," *Mechanical Systems and Signal Processing*, Vol. 18, No. 3, pp. 625~644.

Shao, Y., Chang, C. -H., 2005, "Wavelet Transform to Hybrid Support Vector Machine and Hidden Markov Model for Speech Recognition," *IEEE ISCAS 2005*, pp. 3833~3836.

Tahk, K. M., Shin, K. H., 2002, "A Study on the Fault Diagnosis of Roller-Shape Using Frequency Analysis of Tension Signals and Artificial Neural Networks Based Approach in a Web Transport System," *KSME International Journal*, Vol. 16, No. 12, pp. 1604~1612.

Vapnik, V. N., 1999, "An Overview of Statistical Learning Theory," *IEEE Transaction on Neural Networks*, Vol. 10, No. 5, pp. 988~999.

Xu, Y., Ge, M., 2004, "Hidden Markov Model-Base Process Monitoring System," *Journal of Intelligent Manufacturing*, Vol. 15, pp. 337~350.

Yang, B. S., Han, T., Hwang, W. W., 2005, "Fault Diagnosis of Rotating Machinery Based on Multi-class Support Vector Machines," *Journal of Mechanical Science and Technology*, Vol. 19, No. 3, pp. 846~859.

Yuan, S. F., Chu, F. L., 2006, "Support Vector Machines-Based Fault Diagnosis for Turbo-Pump Rotor," *Mechanical Systems and Signal Processing*, Vol. 20, No. 4, pp. 939~952.