Title

Comparison of Damage Classification Between Recursive Bayesian Model Selection and Support Vector Machine

ISBN
9783319152233

Authors
Mao, Zhu
Todd, Michael

Publication Date
2015

DOI
10.1007/978-3-319-15224-0_11

Peer reviewed

Comparison of Damage Classification between Recursive Bayesian Model Selection and Support Vector Machine

Zhu Mao and Michael Todd

Department of Structural Engineering, University of California San Diego, La Jolla CA 92093-0085; email: mdtodd@ucsd.edu

ABSTRACT

All damage identification activities inevitably involve uncertainties, and the resulting classification ambiguity in contaminated structural health monitoring (SHM) features can dramatically degrade the damage assessment capability. Probabilistic uncertainty quantification (UQ) models characterize the distribution of SHM features as random variables, and the UQ models facilitate making decisions on the occurrence, location, and type of the damages. A Bayesian framework will be adopted and the damage classification is transformed into a model selection process, in which the most plausible structural condition is determined by means of the recursively updated posterior confidence. In contrast to the probabilistic approach, machine learning is another candidate approach, which employs training data and extracts features from the recorded measurements. A support vector machine (SVM) is employed to classify the frequency response function data obtained from rotary machine under different damaged conditions. With different size of feature and different kernel functions, the classification of ball bearing damages are studied. Comparison between the Bayesian model selection approach and SVM is concluded in this paper.

Keywords: Bayesian decision-making, structural health monitoring, damage localization, support vector machine, uncertainty quantification

1. INTRODUCTION

As the fundamental part of structural health monitoring (SHM), a comparison between two system statuses is deployed through the feature domain, and such features are extracted from physical understanding and/or field data acquisition. In reality, this type of decision-making, i.e. distinguishing one state from another, is always corrupted by uncertainties, such as lack of physical intuition, noisy measurements, and environmental/operational variability. To maintain an acceptable quality of SHM decision-making performance, numerous of realizations are often required, and two groups of evidence are compared in a statistical sense naively. To deal with the burden of extensive data acquisition, quantifying the uncertainty in the SHM feature evaluations is necessary. Thereby, the confidence of decision is described through the probabilistic uncertainty quantification (UQ) model, and the overall performance of SHM is enhanced [1-4].

Transfer function, also known as frequency response function (FRF), is one of the most widely-used features for SHM, for the clear physical interpretation and easy-accessibility. UQ models of different estimation algorithms regarding FRF features are established in our previous research, in which probability density functions of the estimates are derived analytically [4,5]. By adopting the probabilistic UQ models, the confidence interval of decision boundaries are pre-defined, and all the testing samples falling outside of the boundaries are labeled as outliers. The percentage of outliers indicates if or not the testing statistics deviates from the undamaged baseline, thus detects damage occurrence.

Moreover, when the statistics of damaged conditions are also known beforehand, a damage classification can be deployed via selecting the most plausible probabilistic model. Specifically in this paper, a Bayesian framework is adopted, because the algorithm fuses the collected evidence to update prior confidence and thereby select the optimal model to characterize the data observation collected from unknown system condition.

As mentioned above, the uncertainty involved in SHM processes causes a lot of burden, especially when extracting sensitive and specific features from large volume of data set. Oftentimes, the great fuzziness and redundancy in the raw data encourage people to investigate powerful feature extraction methodologies. In the past decades, machine learning technologies have been widely applied to SHM, among which support vector machine (SVM) is particularly powerful for solving classification problems. Compared to the Bayesian model selection approach for damage classification, this paper adopts SVM to classify damage cases from the same test-bed.

A brief introduction of the SpectraQuest MFS vibration simulator system is given in section-2, as well as a brief review of the FRF and the UQ model of its estimations. Bayesian model selection approach for classifying damage types will be given in section-3, and SVM implementation, with a parametric study, is available in section-4. In the end, a summary and comparison of the two approaches is given in section-5.

2. TEST-BED AND UNCERTAINTY QUANTIFICATION OF FRF ESTIMATIONS

The SpectraQuest MFS vibration system is adopted as the test-bed to compare the damage classification approaches, as Figure 1 shows. In the simulator system, the bearing on right-hand side of the shaft is altered from undamaged bearing to damaged bearing with defected balls and defected outer race. Acceleration data in direction y and z are recorded, as denoted in Figure 1, and the transfer function between the responses of those two directions are adopted as the damage index.

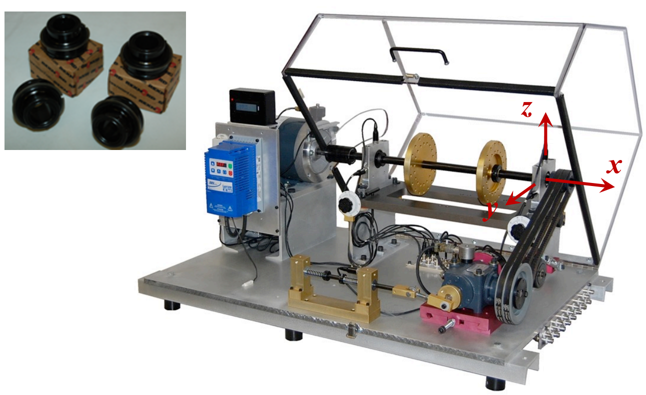


Fig. 1 Rotary machine test-bed

As the ratio between to power spectra, the definition of FRF is described in Equation (1):

$$H(\omega) = \frac{V(\omega)}{U(\omega)}, \tag{1}$$

in which U and V are the Fourier transforms of theoretical input and output $u(t)$ and $v(t)$. When the measurements are contaminated by noise (uncertainty), the realistic input and output are denoted as $x(t)$ and $y(t)$, and the estimation of FRF is often calculated via certainty algorithms, called estimators. Equation (2) is the H-1 estimator of FRF, which is the ratio between cross- and auto-power density functions of (contaminated) input and output measurements:

$$\hat{H}(\omega) = \frac{\hat{G}_{xy}(\omega)}{\hat{G}_{xx}(\omega)}, \tag{2}$$

where the ^ denotes the average of power spectra according to Welch's algorithm [7].

Figure 2 illustrates the FRF feature estimations for various damage conditions, both the magnitude and phase. There are undamaged baseline and other two types of damaged conditions included, as indicated by the three colors in Figure 2. Obviously, the realizations of FRF are very noisy and overlapped at most of the frequency bins. Without investigating the randomness of estimations, it is hardly to make any valuable group classification judgments.

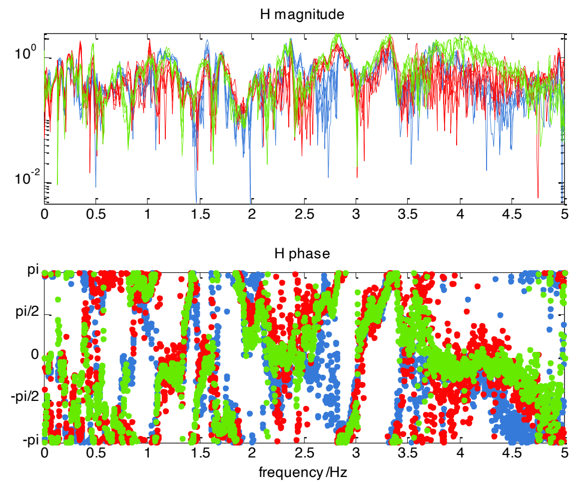


Fig. 2 FRF magnitude and phase estimations for baseline and two damaged conditions

In the Welch's algorithm, the power spectra are estimated in an averaged fashion. If the number of averages is sufficient, the Gaussian distribution is hold asymptotically. The probability density functions of magnitude and phase estimations, as random variables, are derived in [8]:

$$p_m\left(z \mid \mathcal{M}_j\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\left(z-\mu_{m_j}\right)^2}{2\sigma_{m_j}^2}}, \tag{3}$$

where in the context of group classification, $\mathcal{M}_j$ is the jth condition of the structure, and μ_m and σ_m are the mean and standard deviation of magnitude estimation respectively.

The probability density function of phase estimation p_θ is:

$$p_\theta\left(z \mid \mathcal{M}_j\right) = \frac{1}{2\pi} e^{-\frac{\mu_{R_j}^2+\mu_{I_j}^2}{2\sigma_{\theta_j}{}^2}} + \frac{\eta_j}{2\sqrt{2\pi}\sigma_{\theta_j}} \cdot e^{-\frac{\left(\mu_{R_j}\sin(z)-\mu_{I_j}\cos(z)\right)^2}{2\sigma_{\theta_j}{}^2}} \left(1+\operatorname{erf}\left(\frac{\eta_j}{\sqrt{2}\sigma_{\theta_j}}\right)\right), \tag{4}$$

in which $\eta_j = \mu_{R_j}\cos(z) + \mu_{I_j}\sin(z)$, and erf(*) is error function. In Equation (4), μ_{Rj} and μ_{Ij} represent the mean of real and imaginary parts of FRF estimation, while $\sigma_{\theta j}$ is the standard deviation of both parts.

3. BAYESIAN MODEL SELECTION FOR DAMAGE CLASSIFICATION

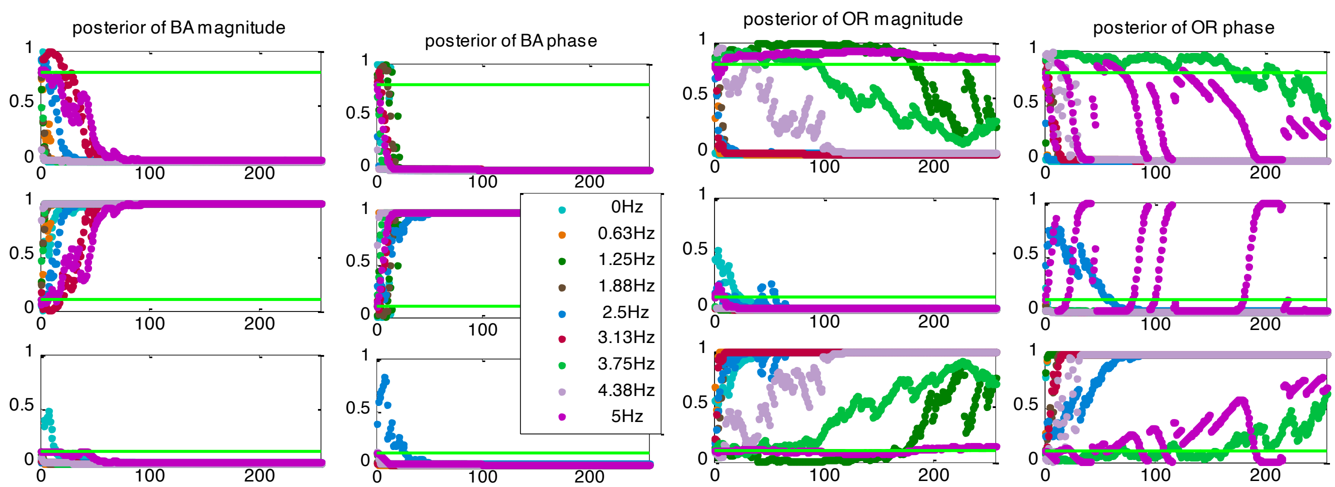
Equation (5) demonstrates the Bayesian framework, which embeds prior knowledge into decision making, and updates the decision confidence when new data are available. The posterior probability $p\left(\mathcal{M}_j \mid \mathcal{D}, \mathbf{M}\right)$ of model (damage condition) $\mathcal{M}_j$ according to Bayes theorem:

$$p\left(\mathcal{M}_j \mid \mathcal{D},\mathbf{M}\right)=\frac{p\left(\mathcal{D}\mid \mathcal{M}_j\right)p\left(\mathcal{M}_j \mid \mathbf{M}\right)}{p\left(\mathcal{D}\mid \mathbf{M}\right)}, \tag{5}$$

in which $p\left(\mathcal{M}_j \mid \mathbf{M}\right)$ is the prior. Likelihood function $p\left(\mathcal{D}\mid \mathcal{M}_j\right)$ is actually the PDF in Equation (3) and (4). Evaluating the total probability theorem, the total evidence for dataset $\mathcal{D}$ on the denominator can be calculated as Equation (6):

$$p\left(\mathcal{D}\mid \mathbf{M}\right)=\sum_{j=1}^{n} p\left(\mathcal{D}\mid \mathcal{M}_j\right)\cdot p\left(\mathcal{M}_j \mid \mathbf{M}\right), \tag{6}$$

where n is the dimension of model class **M**. After running Equation (6) in a recursive fashion for sufficient iterations, i.e. the posterior probability serves as the prior in the next iteration, and posterior updates are produced as the dataset is increased, the posterior $p\left(\mathcal{M}_j \mid \mathcal{D},\mathbf{M}\right)$ will tend towards 1 or 0, indicating acceptance or rejection of the jth model (damage condition).



(a) Using FRF magnitude feature, damaged ball data | (b) Using FRF phase feature, damaged ball data | (c) Using FRF magnitude feature, damaged race data | (d) Using FRF phase feature, damaged race data

Fig. 3 Recursive posterior probability of model selection, at sample frequencies; top: baseline model; middle: damaged ball model; bottom: damaged outer race model

In each class, the posterior probability of selecting among baseline (top), ball defect (middle), and race defect (bottom), as a function of number of iterations is plotted, and each color indicates a different frequency line. It is clear that in all cases, the posterior converges to the right number, i.e. converging to 1 if there is damage, and converging to 0 if not. The horizontal lines in green highlight the arbitrarily-picked prior probability before any testing information. In Figure 3(c) and 3(d), the convergence is not as decisive as Figure 3(a) and 3(b). That illustrates for damaged race, not all the sampled frequencies has the same detection capability

4. SUPPORT VECTOR MACHINE FOR DAMAGE CLASSIFICATION

4.1 SVM and kernelization

Different from the probabilistic approach using Bayesian framework, support vector machine (SVM) employs training data to form a hyperplane as the decision boundaries, in order to discriminate different sets of data. All the data points determining the hyperplane are called support vectors. Equation (5) describes the decision maker h(*), which maps feature vector x into a binary space:

$$\mathrm{h}(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathrm{g}(\mathbf{x}) > 0 \\ 1 & \text{if } \mathrm{g}(\mathbf{x}) < 0 \end{cases}, \tag{7}$$

The function g(*) forms a hyperplane in the feature domain, which is described as:

$$\mathrm{g}(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + \mathbf{b} \cdot \tag{8}$$

For the realistic data presented in last section, the data from the classification model are mostly non-separable, thus the slack variables may be introduced to solve a soft margin problem. It is not always practical for highly overlapped/complicated feature spaces. Kernel functions are employed if necessary, to introduce extra feature dimensions, and all clusters are being better distinguished in a higher dimensional state, as Equation (9) shows:

$$\mathrm{h}(\mathbf{x}) = \mathrm{sign}\left[\sum_{i \in \mathrm{SV}} y_i \boldsymbol{\alpha}_i^* \mathrm{K}(\mathbf{x}_i, \mathbf{x}) + \mathbf{b}^*\right]. \tag{9}$$

in which $\mathbf{x}_i$ are all the support vectors, K(*,*) is the selected kernel function, $\boldsymbol{\alpha}_i$ is the Lagrange multiplier for constrained optimization, and y_i is the classification label of feature $\mathbf{x}_i$.

4.2 SVM classification

Speaking of the cluster of features, a simplified condition is considered at the beginning for SVM analysis. Only a single frequency line is considered and the feature vector is defined as:

$$\mathbf{x} = \left[|H(\omega)| \quad \measuredangle H(\omega) \right]^T \in \sim^{2}. \tag{10}$$

In Equation (10), ω is a sample frequency, and apparently the feature $\mathbf{x}$ is a 2-D vector. Figure 4 visualizes the feature vector, and the overlap of clusters is obvious.

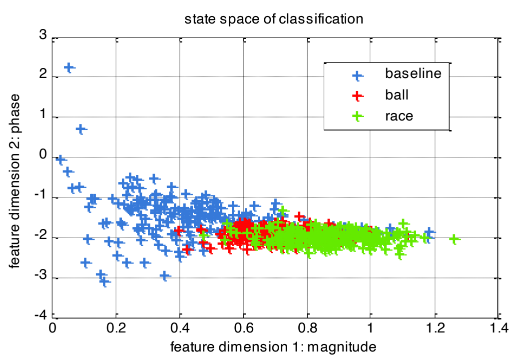


Fig. 4 Feature state space at a sample frequency line

In the context of SHM, SVM provides the boundary to separate damaged data from undamaged baseline. Figure 5 and 6 demonstrate such damage detection implementation on a binary case, with only ball damage involved. With about 250 testing cases in total, the classification result and the true condition for each test is plotted on the right. For each class, the percentage of correct labelling from SVM is calculated, for comparison with the random guess rate. In addition, the kernel trick in SVM increases the decision space dimensionality, so that the decision boundaries plotted are actually the projection of the real decision boundaries to the 2-D plane. Two different kernel realizations are adopted, namely linear kernel and Gaussian kernel, which is a more flexible radial bases function.

Comparing the rate of correct classification marked in the figures, the Gaussian kernel does better than the linear hyperplane separation. However, the performance of correct classification rate has to be compared considering all classes. For instance, if the classification is totally flipped (wrong), the binary classification rate will be 0%/0%, and 50%/50% will indicate an ideal random guess, and only 100%/100% means the perfect classification. The case of 0%/100% does not suggest a good classification, because in this scenario, the algorithm just classifies all the test cases into one class.

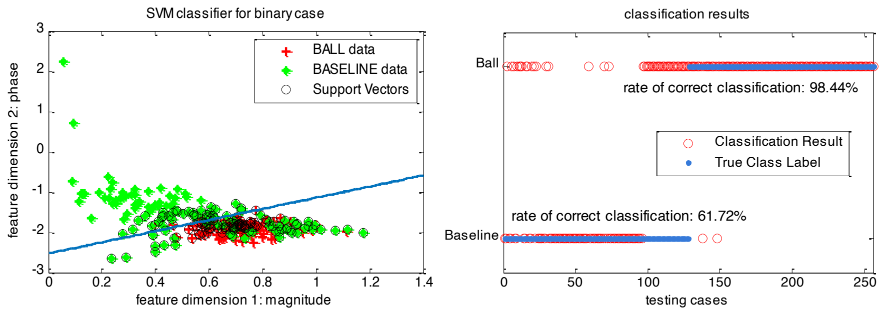


Fig. 5 SVM implementation for binary classification, linear kernel

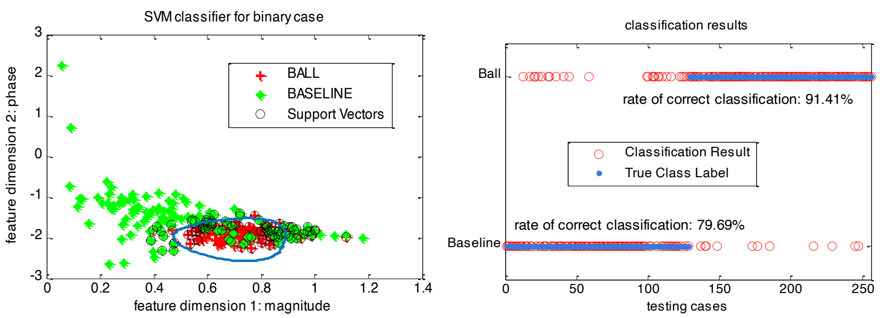


Fig. 6 SVM implementation for binary classification, Gaussian kernel

For the trinary classification problem, two damaged conditions need to be discriminated from the baseline, and this multi-class classification is implemented into three binary classifications. Each of the binary sub-problems is a "distinguishing one from all the others" approach, which is essentially the same as damage detection. Specifically speaking, the one-versus-rest idea partitions the trinary classification into (1 vs 2, 3), (2 vs 3, 1) and (3 vs 1, 2), all three sub-problems.

Because the Gaussian kernel outperforms slightly according to Figure 5 and 6, the multi-class discriminations presented in Figure 7 for each sub-problem employ Gaussian kernel. Just for a comparison, the result of "correct labeling" is also available in Figure 8. In each sub-problem, the class to be distinguished is highlighted and the other two classes are grouped as "ELSE".

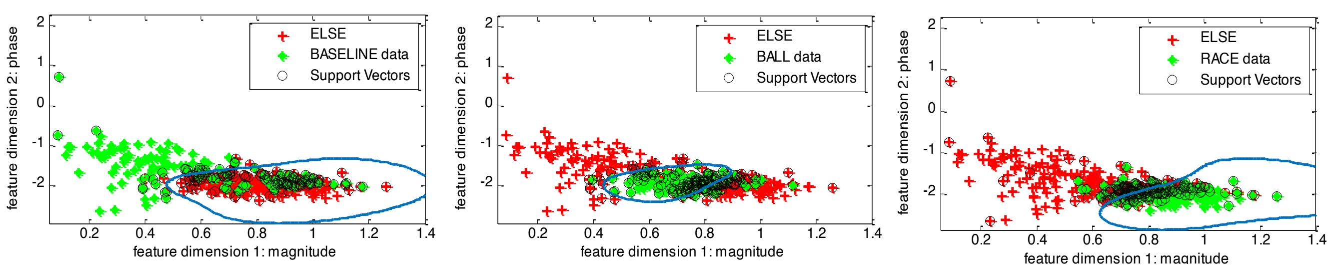


Fig. 7 SVM implementation for trinary classification, Gaussian kernel

The same as damage detection, the performance of SVM for trinary classification is evaluated via correct rate, plotted in Figure 8. The same conclusion can be made, that the Gaussian function outperforms linear function due to the flexibility characteristics. Compared to the binary case, the SVM classifiers obtained from both kernel functions, in general, tag the data less accurate than the performance of binary case. This is mainly caused by the heavier complexity and ambiguity. The average rate of correct classification for each case is about 57% and 71%, which is lower than the binary classification, given the same feature dimension.

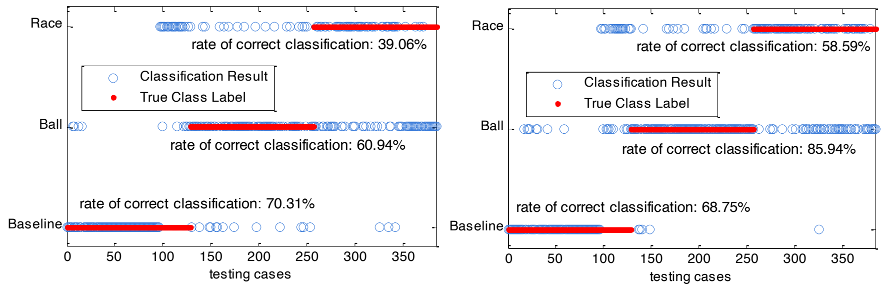


Fig. 8 Correct classification rate for SVM implementation for trinary classification
left: linear kernel; right: Gaussian kernel

So far, all the damage detection/classification are implemented based upon the FRF feature at single frequency, and the correct labeling rate for binary and trinary is 85% and 71% respectively, using Gaussian kernel. In fact, a lot of information has not been utilized, because most of the spectral characteristics are thrown away when selecting the sample frequency. Instead of eliminating most of the FRF information, Equation (11) defines the state space feature by using more frequency lines:

$$\mathbf{x} = \left[\left|H(\omega_1)\right|,\ \left|H(\omega_2)\right| \ \cdots \ \left|H(\omega_n)\right|,\ \measuredangle H(\omega_1),\ \measuredangle H(\omega_2) \ \cdots \ \measuredangle H(\omega_n)\right]^T \in \sim^{2n}. \tag{11}$$

in which n frequencies are considered to build up the new feature vector $\mathbf{x}$.

The new-established feature state has a lot more dimension. For example, if all the frequency lines are included in the vector $\mathbf{x}$, there will be around a thousand state space dimensions. Under this circumstance, the state space is not able to be visualized, and Figure 9 only plots the projection to three arbitrary dimensions, which is the limit on a 2-D plane.

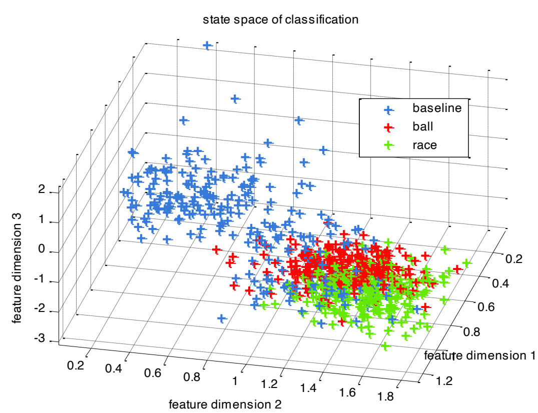


Fig. 9 High-dimension feature state space projected to arbitrary three dimensions

Given feature space with ultra-high dimension, Gaussian kernelization becomes burdensome and could not return a reliable classifier easily, especially due to the relative lack of training data. As a result, linear kernel function is adopted, for both binary and trinary classification. By using the high-dimensional feature, the rates of correct classification for both binary and trinary cases are 100%, as shown in Figure 10, which means all the test cases are correctly labeled.

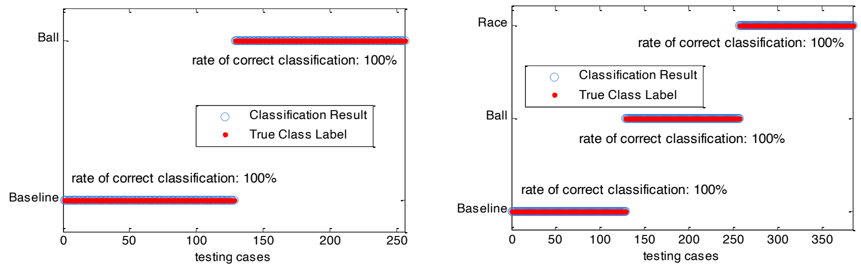


Fig. 10 Rate of correct classification via SVM classifiers, linear kernel
left: binary classification; right: trinary classification

5. SUMMARY AND CONCLUSION

This paper compares two strategies of damage detection and classification, namely a Bayesian model selection approach and support vector machine classification. Bearing defects on rotary machine is selected as the test-bed to implement the strategies. In the Bayesian approach, it takes tens of samples for the posterior probability of selecting/denying a model to saturate to either 1 or 0, but for SVM, the training and testing will take longer time. In SVM approach, the performance for binary classification is better than for trinary cases, because of the lower complexity, and higher-dimensional features will lead to a more specific classification. On the other hand, higher-dimensional feature spaces will cause more computational burden in the SVM training procedure, which is a major drawback compared to Bayesian approach. Among the two types of kernel functions, SVM with linear kernel function separates data with a hyperplane and has the advantage of relatively faster training, while Gaussian kernel has better flexibility to handle more complicated data sets. This gets verified by high-dimensional features, where Gaussian kernel may not converge to a decision boundary with limited training sets. Although the SVM is computational expensive, compared to the Bayesian approach, it is better at handling high-dimensional features, because when the feature space expands, the joint probability density function will have an exponentially increasing complexity, and the evaluation of likelihood functions will become troublesome.

ACKNOWLEDGEMENT

This research was supported by the research grant (UD130058JD) of the Agency for Defense Development of the Korean government and by the Leading Foreign Research Institute Recruitment Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT and Future Planning (2011-0030065).

Reference

[1] J.S. Bendat, A.G. Piersol. Random Data: Analysis and Measurement Procedures (2nd Edition). John Wiley & Sons (New York), 1986.
[2] D.R. Brillinger, Time Series: Data Analysis and Theory. McGraw-Hill, New York, 1981.
[3] B.R. Mace, K. Worden, G. Manson, Uncertainty in structural dynamics. *Journal of Sound & Vibration* 288 423-429, 2005.
[4] Z. Mao, M.D. Todd. A model for quantifying uncertainty in the estimation of noise-contaminated measurements of transmissibility. *Mechanical Systems and Signal Processing*. Volume 28, p. 470-481, 2012.
[5] Z. Mao, M.D. Todd. Statistical modeling of frequency response function estimation for uncertainty quantification. *Mechanical Systems and Signal Processing*. Volume 38, p.333–345, 2013.
[6] Z. Mao, M. D. Todd. Optimal Structural Health Monitoring Feature Selection via Minimized Performance Uncertainty. *Key Engineering Materials*. Vol. 558, p.235-243, 2013.
[7] P.D. Welch. The use of Fast Fourier Transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoustics*. 15(70-73), 1967.
[8] Z. Mao, M. D. Todd. Bayesian Updating of Detection Capability with Frequency Response Function Related Structural Health Monitoring Features. *the Proceedings of the 9th International Workshop on Structural Health Monitoring* (IWSHM 2013)