Active Learning Data Selection for Adaptive Online Structural Damage Estimation

Debejyo Chakraborty†, Narayan Kovvali†, Antonia Papandreou-Suppappola†, and Aditi Chattopadhyay‡

†School of Electrical, Computer & Energy Engineering
‡School of Mechanical, Aerospace, Chemical & Materials Engineering
Arizona State University, Tempe, AZ 85287

ABSTRACT

Adaptive learning techniques have recently been considered for structural health monitoring applications due to their flexibility and effectiveness in addressing real-world challenges such as variability in the monitoring of environmental and operating conditions. In this paper, an active learning data selection procedure is proposed that adaptively selects the most informative measurements to include, from multiple available measurements, in estimating structural damage. This is important, since not all the measurements may provide useful information and could reduce performance when processed. Within the adaptive learning framework, the data selection problem is formulated to choose those measurements which are most representative of the diversity within a damage state. This is achieved by extracting time-frequency analysis based statistical similarity features from the measurements, and selecting uniformly distributed subsets to build representative reference sets. The utility of the proposed method is demonstrated by improvements in adaptive learning performance for the estimation of fatigue damage in an aluminum compact tension sample.

Keywords: structural health monitoring, damage classification, adaptive learning, active data selection, time-frequency analysis, matching pursuit decomposition, statistical similarity, optimization, discrepancy

1. INTRODUCTION

Several structural health monitoring (SHM) techniques are lately being developed using advanced material studies, signal processing and statistical methods.[1–4] However, they do not address variability in operating condition like changes in loading, in the environment like temperature or pressure, or in the material composition which affect the system health.

Adaptive learning techniques[5–7] are state-of-the-art methods that are attractive due to their flexibility and effectiveness in real-world learning scenarios. We have recently considered[8] these methods in SHM[2] applications where variability in environmental and operating conditions is a big challenge. In our proposed adaptive learning methodology in,[8] we adopted a state-space formulation that describes the evolution of damage state by a Markov relationship that accounts for variability in the environment or the structure itself. The observations in the formulation are composed of clusters identified adaptively using Dirichlet process (DP) mixture models[9,10] of time-frequency (TF)[11] features that are extracted from the sensor measurements. Once the formulation is established, Bayesian filtering[12] can be used to estimate the progressive damage.

The block diagram of our adaptive learning based damage state estimation method in[8] is shown in Figure 1. The extraction of a reference time-frequency representation (TFR) subset from the adaptive clustering results is shown to only use the data from the most recent epoch, and as such may not be representative of the variability in the data. This approach does not exploit the fact that an optimal choice of the data is expected to yield maximum damage information.

Active data selection is a powerful approach for intelligently selecting the most informative measurements, and has found application in various fields of engineering.[13–16] In the context of the adaptive learning SHM

Further author information: (Send correspondence to Debejyo Chakraborty)
E-mail: Debejyo.Chakraborty@asu.edu, Web: http://www.debejyo.com

Nondestructive Characterization for Composite Materials, Aerospace Engineering, Civil Infrastructure, and Homeland Security 2010, edited by Peter J. Shull, Aaron A. Diaz, H. Felix Wu, Proc. of SPIE Vol. 7649
764915 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.848891

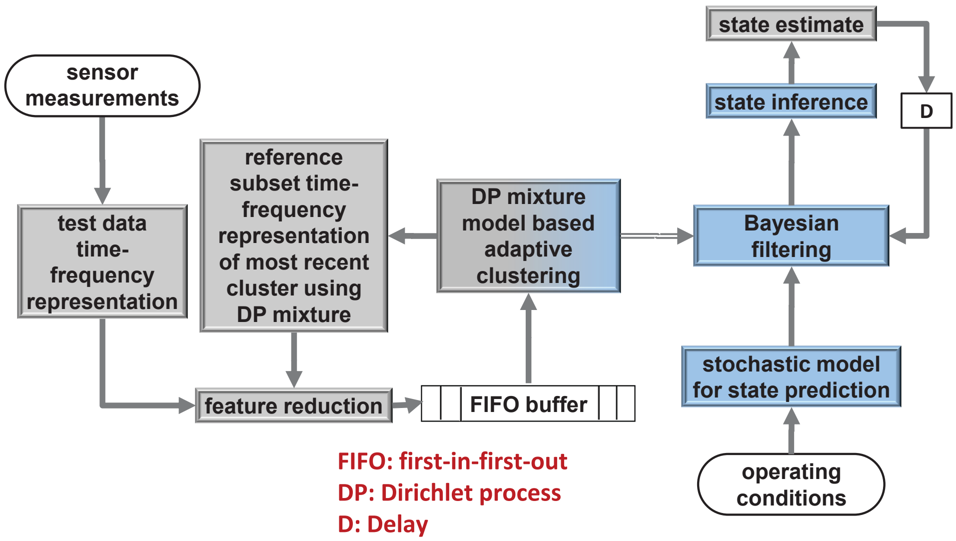


Figure 1. Block diagram of the adaptive learning based damage state estimation method.[8]

framework, active data selection provides a mechanism to address the variability inherent in the measurements. For example, it is possible that some of the measurements came from the defective sensors or that not all the measurements considered demonstrate a large spread in diversity for the given damage state.

In this paper, we propose an active learning data selection scheme that automatically selects the measurements which are most representative of the diversity within a damage state. TF analysis based statistical similarity features are first extracted from the measurements using the matching pursuit decomposition (MPD) algorithm.[17] MPD based TF features have been shown[4,18–22] to provide highly localized characteristics of the time-varying spectral nature of damage wave-physics. Here, we condense the TF information contained in the measurements into scalar statistical similarity features. Active data selection is then used to design the reference sets in the adaptive learning SHM framework. In particular, uniformly distributed samples are selected from the available data by optimizing a discrepancy uniformity measure.[23,24] We demonstrate the proposed method by improvements in adaptive learning performance for the estimation of fatigue damage in an aluminum compact tension sample.

The remainder of this paper is organized as follows. In Section 2, we introduce the theoretical framework of the proposed feature extraction and data selection method. Section 3 presents an application of the proposed method to the adaptive learning used in fatigue damage estimation.

2. THEORETICAL DEVELOPMENT OF DATA SELECTION ALGORITHM

In this section, we briefly describe the theoretical framework of the proposed feature extraction and data selection method. More details can be found in the literature.[11,17,23–26]

2.1 Measurement Probability Density Function

Considering the adaptive learning based damage state estimation method in Figure 1, we first develop different probabilistic models for the time-frequency features extracted from the sensor measurements of different damage states using the matching pursuit decomposition (MPD) algorithm. The MPD[17] is an iterative method for representing a signal by a linear combination of basis functions chosen from a redundant dictionary. Specifically, a signal $s(t) \in L^2(\mathbb{R})$ is approximated as

$$s(t) \approx \sum_{n=0}^{N-1} \alpha_n \, g_n(t), \tag{1}$$

where the functions $g_n(t)$ are chosen iteratively from a dictionary. The expansion coefficients α_n are projections of $s(t)$ on to $g_n(t)$. The accuracy of the approximation improves with increasing basis functions N.

A popular choice of dictionary functions is provided by the Gaussian which have the important property of achieving maximum time-frequency localization.[11] For an MPD with Gaussian dictionary, the dictionary elements are given by the Gaussian function

$$g^{(d)}(t) = \left(\frac{2\kappa_r}{\pi}\right)^{\frac{1}{4}} \exp\left(-\kappa_r(t-\tau_m)^2\right) \exp(j2\pi\nu_l t), \tag{2}$$

where $d = \{\tau_m, \nu_l, \kappa_r\}$, τ_m is the mth time shift, ν_l is the lth frequency shift, κ_r is the rth scale change, and the range values of m, l, and r are such that the TF plane spans the maximum TF support of the analysis signals. The Wigner distribution (WD) of the transformed Gaussian dictionary element can be computed is closed form[11] as

$$\mathrm{WD}_{g^{(d)}}(t, f) = 2\exp\left(-2\kappa_r(t-\tau_m)^2\right) \exp\left(-\frac{2\pi^2(f-\nu_l)^2}{\kappa_r}\right). \tag{3}$$

The MPD time-frequency representation (MPD-TFR)[17] of the signal $s(t)$ in (1) is

$$\mathcal{E}_s(t, f) = \sum_{n=0}^{N-1} |\alpha_n|^2 \mathrm{WD}_{g_n}(t, f). \tag{4}$$

We emphasize that, at the nth iteration, the MPD chooses the Gaussian function $g_n(t)$ with corresponding transformation parameter features $(\tau_n, \nu_n, \kappa_n)$ whose WD is the two-dimensional (2-D) Gaussian function in Equation (3). The 2-D Gaussian function WD can be interpreted as a 2-D Gaussian probability density function with mean vector $\begin{bmatrix} \tau_n \\ \nu_n \end{bmatrix}$ and covariance matrix $\begin{bmatrix} \frac{1}{4\kappa_n} & 0 \\ 0 & \frac{\kappa_n}{4\pi^2} \end{bmatrix}$. Thus, we will use the 2-D Gaussian probability density function interpretation for the stochastic measurement model. We define the MPD based probability density function (MPD-PDF) for the measurement $s(t)$ as

$$P_s(t, f) \triangleq \frac{1}{Z} \left|\mathcal{E}_s(t, f)\right|, \tag{5}$$

where $Z = \sum_{n=0}^{N-1} |\alpha_n|^2$ is the normalizing constant. For a dictionary of Gaussian functions, the MPD-PDF assumes the form of a Gaussian mixture model (GMM) defined as

$$P_s(t, f) = \frac{1}{Z} \sum_{n=0}^{N-1} |\alpha_n|^2 \mathcal{N}\left(\begin{bmatrix} \tau_n \\ \nu_n \end{bmatrix}, \begin{bmatrix} \frac{1}{4\kappa_n} & 0 \\ 0 & \frac{\kappa_n}{4\pi^2} \end{bmatrix}\right), \tag{6}$$

where $\mathcal{N}(\cdot, \cdot)$ is a Gaussian distribution with a mean and a covariance.

The MPD-PDF maps signals to two-dimensional (time-frequency) probability density functions and provides a versatile tool for computing statistical similarity measures between sensor measurements.

2.2 Statistical Similarity Measures and Signal Classification

The statistical similarity of two PDFs can be quantified by well known measures such as Kullback-Leibler divergence (KLD),[25,27,28] Bhattacharyya distance,[29,30] and Hellinger distance.[31] Among these, the most popular is the KLD defined for PDFs $p(t, f)$ and $q(t, f)$ as*

$$\mathrm{D}^{\mathrm{KL}}(p||q) \triangleq \iint p(t, f) \log \frac{p(t, f)}{q(t, f)} \, dt \, df. \tag{7}$$

*Unless otherwise stated, limits of integration ranges from $-\infty$ to $+\infty$ throughout the paper.

Note that $\mathrm{D}^{\mathrm{KL}}(p||q) \geq 0$, with equality if and only if $p(t,f) = q(t,f)$. However, this distance is non-symmetric, i.e., $\mathrm{D}^{\mathrm{KL}}(p||q) \neq \mathrm{D}^{\mathrm{KL}}(q||p)$, and its computation involves evaluating logarithms, which is expensive. In,[8] a correlation based distance

$$\mathrm{D}^{\mathrm{corr}}(p||q) \triangleq \iint p(t,f)\, q(t,f)\, dt\, df \tag{8}$$

is considered which satisfies $\mathrm{D}^{\mathrm{corr}}(p||q) \geq 0$. The distance measure can be shown to be maximum when $p(t,f) = q(t,f)$ and minimum when the two densities are orthogonal. This distance is also symmetric and can be computed efficiently using Monte Carlo integration as described next.

Let $\{t_l, f_l\}_{l=1}^{L} \sim p(t,f)$ denote independent and identically distributed (i.i.d.) samples drawn from $p(t,f)$. Then, the distance $\mathrm{D}^{\mathrm{corr}}(p||q)$ in (8) can be approximated[32] by

$$\hat{\mathrm{D}}^{\mathrm{corr}}(p||q) = \frac{1}{L} \sum_{l=1}^{L} q(t_l, f_l), \tag{9}$$

with the accuracy of the approximation increasing with the number of samples L.

The statistical similarity measures can be used as features for signal classification. Specifically, given two signals $s_1(t)$ and $s_2(t)$, we can investigate their similarity by first computing the respective MPD-PDFs, $P_{s_1}(t,f)$ and $P_{s_2}(t,f)$ using Equation (6), and then evaluating, for example, $\mathrm{D}^{\mathrm{corr}}(P_{s_1}||P_{s_2})$. Similarly, if we wish to find the similarity between a given test signal $s^{\mathrm{test}}(t)$ and a set of Reference signals $\{s_1^{\mathrm{ref}}(t), \ldots, s_R^{\mathrm{ref}}(t)\}$, we can evaluate

$$\mathrm{D}^{\mathrm{avg}} = \frac{1}{R} \sum_{r=1}^{R} \mathrm{D}^{\mathrm{corr}}(P_{s^{\mathrm{test}}}||P_{s_r^{\mathrm{ref}}}) = \mathrm{D}^{\mathrm{corr}}(P_{s^{\mathrm{test}}}||P_{s^{\mathrm{ref}}}), \tag{10}$$

where $P_{s^{\mathrm{ref}}}(t,f) = \frac{1}{R}\sum_{r=1}^{R} P_{s_r^{\mathrm{ref}}}(t,f)$ is the average MPD-PDF for the reference set. The PDF $P_{s^{\mathrm{ref}}}(t,f)$ collects the statistical variability in the reference set, and is a GMM in the Gaussian dictionary framework described in Section 2.1.

In the classical training-testing based classification approach, multiple signals are measured to ensure that the statistical variability in the reference class is captured. However, most classification problems stand to benefit from judicious choice of the reference set. In particular, sets with statistical diversity are preferred.

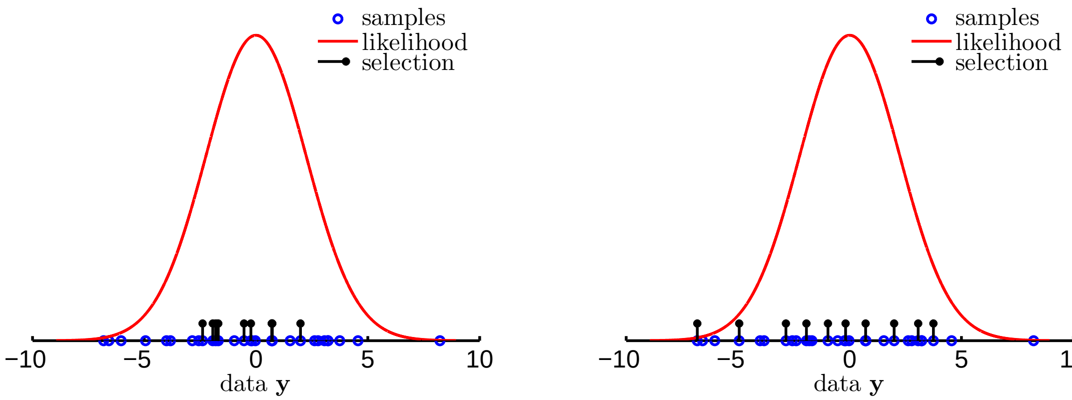
2.3 Uniform Data Selection

The goal of uniform data selection is to choose a subset of *uniformly distributed* points from a given sample set. Figure 2 shows a simulated example from a data set comprising samples from a Gaussian distribution. The samples selected in Figure 2(a) are clustered near the mean of the Gaussian distribution (non-uniform) whereas those in Figure 2(b) are distributed more uniformly over the data space and are therefore more representative of the entire data space.

The notion of uniformity can be quantified by the *discrepancy*[23] $\mathcal{D}_M(\mathbf{y})$ defined for a set of points $\mathbf{y} = \{y_1, \ldots, y_M\}$ with respect to an interval $[y^{\min}, y^{\max}]$ as

$$\mathcal{D}_M(\mathbf{y}) \triangleq \sup_{y^{\min} \leq y^l \leq y^h \leq y^{\max}} \left| \frac{\left|\mathbf{y} \cap [y^l, y^h]\right|}{M} - \frac{y^h - y^l}{y^{\max} - y^{\min}} \right|, \tag{11}$$

where $[y^l, y^h]$ defines any subinterval of $[y^{\min}, y^{\max}]$ and $\left|\mathbf{y} \cap [y^l, y^h]\right|$ represents the cardinality of $\mathbf{y}$ in $[y^l, y^h]$. The set $\mathbf{y}$ is said to be uniformly distributed on the interval $[y^{\min}, y^{\max}]$ if $\lim_{M\to\infty} \mathcal{D}_M(\mathbf{y}) = 0$.[23,24] For a given set of points $\mathbf{y} = \{y_1, \ldots, y_M\}$ in $[y^{\min}, y^{\max}]$, the discrepancy $\mathcal{D}_M(\mathbf{y})$ can be used as a measure of non-uniformity. In particular, the discrepancy is observed to be minimum for a perfectly uniform $\mathbf{y}$ in $[y^{\min}, y^{\max}]$. In the example shown in Figure 2, the discrepancy of the selected samples is greater in Figure 2(a) than it is in Figure 2(b).



(a) Non-uniform data selection (b) Uniform data selection

Figure 2. A comparison of non-uniform vs. uniform data selection.

The uniform data selection method proposed in this paper is based on the discrepancy of the sensor measurements. Specifically, we define an indicator vector $\mathbf{z} = \{z_1, \ldots, z_M\}$ to indicate the samples selected from the measurement $\mathbf{y}$:

$$z_i = \begin{cases} 1, & \text{if } y_i \text{ is selected} \\ 0, & \text{otherwise} \end{cases}, \; i = 1, \ldots, M.$$

Let K be the number of samples desired in the selected measurements subset $\mathbf{y}_{\mathbf{z}}^{\text{sel}}$. The uniform data selection problem can now be stated formally as: given a data set comprising of M points $\mathbf{y} = \{y_1, \ldots, y_M\}$ in the interval $[y^{\min}, y^{\max}]$, find a selection $\mathbf{z}^*$ of cardinality K which minimizes the discrepancy $\mathcal{D}_K(\mathbf{y}_{\mathbf{z}}^{\text{sel}})$. The optimum selection is obtained by solving the problem

$$\begin{array}{ll} \text{minimize} & \mathcal{D}_K(\mathbf{y}^{\text{sel}}(\mathbf{z})) \\ \text{subject to} & \mathbf{z}_i \in \{0, 1\}, \;\; i = 1, \ldots, M \\ & \|\mathbf{z}\|_0 = K. \end{array} \tag{12}$$

Here, $\|\mathbf{z}\|_0$ represents the cardinality of the vector $\mathbf{z}$. The minimization problem in (12) is non-convex and in general very difficult to solve exactly.[26] Direct enumeration is ruled out because the number of possible selections is $\binom{M}{K}$ which is very large even for relatively small M and K. Evaluating the discrepancy can also be computationally expensive, and some of the estimation techniques to compute it include techniques based on sampling, searching, lower and upper bounds, and convex programming.[24, 33–35]

In this paper, we propose an approximate and efficient method for uniform data selection by considering the relaxed minimization problem

$$\begin{array}{ll} \text{minimize} & \tilde{\mathcal{D}}_K(\mathbf{y}^{\text{sel}}(\mathbf{z})) \\ \text{subject to} & \mathbf{z}_i \in \{0, 1\}, \;\; i = 1, \ldots, M \\ & \|\mathbf{z}\|_0 \approx K, \end{array} \tag{13}$$

where the objective function is defined as

$$\tilde{\mathcal{D}}_M(\mathbf{y}) \triangleq \frac{1}{J} \sum_{h=1}^{J} \left| \frac{\left|\mathbf{y} \cap [y^{\min} + (h-1)\Delta y, y^{\min} + h\Delta y]\right|}{M} - \frac{\Delta y}{y^{\max} - y^{\min}} \right|, \tag{14}$$

with $\Delta y = (y^{\max} - y^{\min})/J$. Here, J denotes the number of equal length partitions in $[y^{\max} - y^{\min}]$. While the discrepancy in (11) is defined as a supremum over all possible subintervals of $[y^{\min}, y^{\max}]$, the objective function in (14) is defined as an average over the J equal-length subintervals partitioning $[y^{\min}, y^{\max}]$. The constraint of the minimization problem in (14) on the number of data points to be selected has also been relaxed. As we show using simulations, the solution of (13) leads to reasonably uniform data selection.

The problem (13) can be solved efficiently by noting that the objective can be minimized term-by-term. This is achieved by selecting the data such that $\left|\mathbf{y}^{\text{sel}}(\mathbf{z}) \cap [y^{\min} + (j-1)\Delta y, y^{\min} + j\Delta y]\right| \approx \mathbf{round}\,(K/J)$ for $j = 1, \ldots, J$. Here, $\mathbf{round}\,(K/J)$ gives integer value nearest to K/J. Algorithm 1 contains the pseudocode[†] for the solution of (13). The run-time and storage requirements are roughly $\mathcal{O}(K)$ and $\mathcal{O}(M)$, respectively.

Algorithm 1 Uniform data selection via solution of problem (13)

```
z ← 0
n ← round (K/J)
for j = 1 to J do
  I ← {i : y_i ∈ [y^min + (j − 1)Δy, y^min + jΔy]}
  m ← |I|
  if m ≤ n then
    z_I ← 1
  else
    for k = 1 to n do
      Draw u ~ U[0, 1]
      w ← ⌈u · m⌉
      z_I(w) ← 1
      I ← I\I(w)
      m ← m − 1
    end for
  end if
end for
```

3. APPLICATION AND RESULTS

We now present an application of the proposed feature extraction and data selection method for adaptive learning used in fatigue damage estimation.

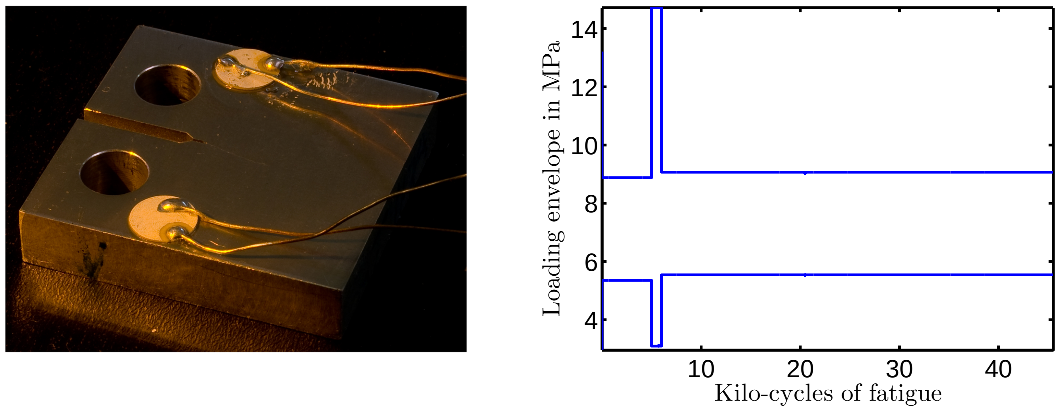
Figure 3(a) shows the aluminum 2024 compact tension (CT) sample used in our experiments. The width of the sample is 25.4 mm. A variable amplitude cyclic load of up to about 45 kilocycles was applied to the sample to induce fatigue crack damage. The amplitude envelope of the applied load is shown in Figure 3(b). Two surface mounted piezoelectric (PZT) sensors (placed symmetrically on the sample as shown in Figure 3(a)) were used to make measurements at several stages of fatigue loading cycles. One PZT was used as the actuator and the other was used as a sensor to measure the response signals. A 130 kHz burst signal was used for excitation. Further details of the experimental setup and data collection can be found in.[36]

After pre-processing the measurements for normalization, MPD was performed on each measurement with $N = 10$ iterations and the MPD-TFRs were constructed. The MPD-PDFs were then defined using (5). Figure 4 shows the MPD-TFRs of two signals measured obtained by successive measurements for a crack length of 6.17 mm. From the plots, we see marked difference in the time-frequency structure of signals from the same damage state. This example demonstrates the inherent variability present in measurements from a given damage state.

Active data selection provides a mechanism to address this variability. In this work, we incorporate the proposed uniform data selection algorithm into the adaptive learning framework described in[8] (see block diagram in Figure 1). Specifically, at any given epoch, we first compute the statistical similarity features between the current measured signals and the reference set (Equation (10)). Adaptive clustering is then performed on these features and used subsequently for damage estimation (see[8] for more details). Finally, the active data selection is used to update the reference set.

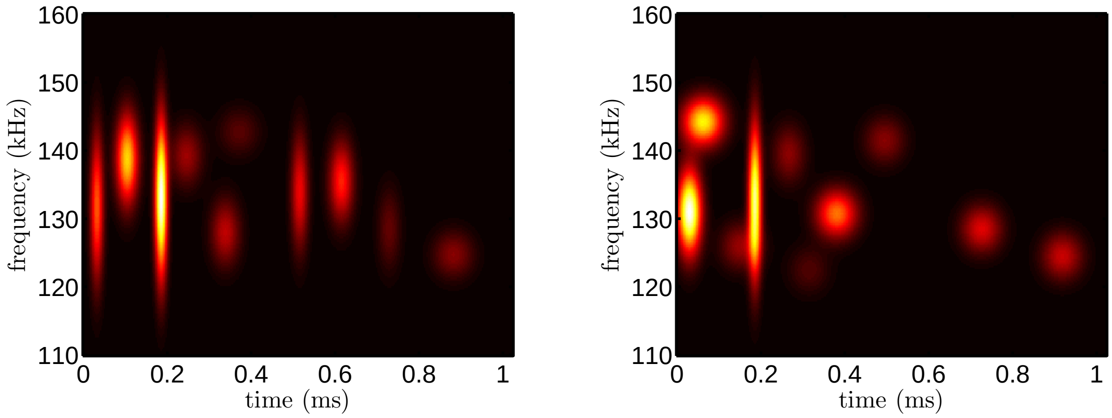
In our simulations, at each epoch, $K = 10$ (out of $M = 30$) samples are selected for the reference set. Figure 5 compares the performance of the adaptive clustering at one epoch with and without active data selection. Note that we obtain similar performance at different time epochs. Figure 5(a) shows the data samples (statistical

[†]Note that $\mathcal{I}(w)$ is used to represent the wth element of set $\mathcal{I}$.



(a) CT sample with actuator, sensor, and crack (b) Amplitude envelope of the applied load

Figure 3. CT sample and loading profile.



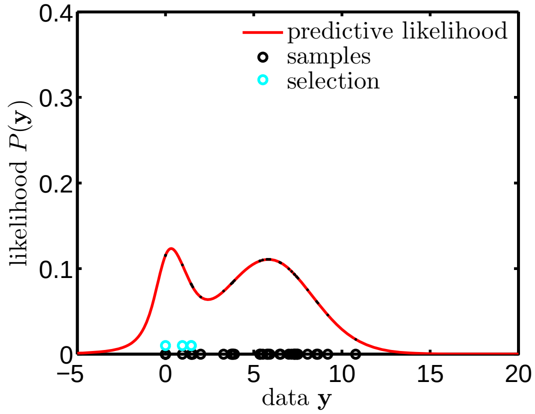
(a) (b)

Figure 4. MPD-TFRs of two signals for a crack length of 6.17 mm.

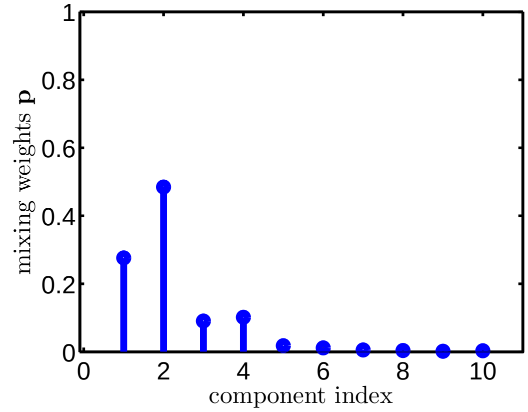
similarity features), the adaptively learned likelihood, and the randomly selected samples for the reference set. The corresponding adaptive clustering results are given in Figure 5(b). Figure 5(c) shows the data samples, the adaptively learned likelihood, and the uniformly selected samples for the reference set. The corresponding adaptive clustering results are given in Figure 5(d). Given that the data in this example belongs to the same damage state, good clustering is expected to identify a single cluster. From the plots, we see that this happens when using uniformly selected data. In the case of randomly selected data, the number of clusters is incorrectly identified as two.

4. CONCLUSION

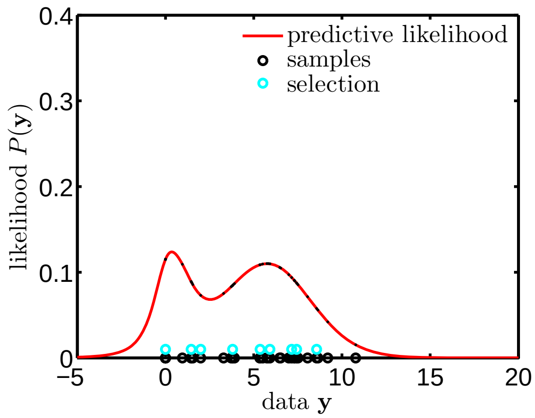
In this paper, we presented an active data selection procedure that is formulated within the adaptive learning framework to choose measurements representative of the diversity within a damage state. Time-frequency analysis based statistical similarity features are extracted from the measurements, and representative reference sets are built using uniformly distributed subsets of data. The proposed feature extraction and data selection method is applied to the adaptive learning used for the estimation of fatigue damage in an aluminum CT sample. Simulations results indicate that the proposed algorithm yields improvements in adaptive clustering performance, increasing the reliability of the damage identification process. The new framework is computationally efficient and easy to implement.



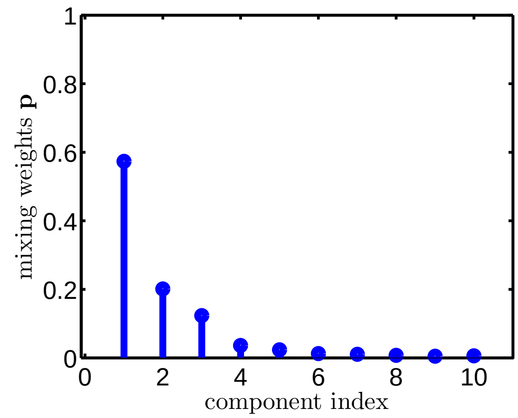
(a) Randomly selected data



(b) Inaccurate clustering



(c) Uniformly selected data



(d) Accurate clustering

Figure 5. Active data selection for improved adaptive learning.

ACKNOWLEDGMENTS

This research was supported by the MURI Program, Air Force Office of Scientific Research, grant number: FA9550-06-1-0309; Technical Monitor, Dr. David Stargel.

REFERENCES

[1] Morassi, A. and Vestroni, F., eds., [*Dynamic Methods for Damage Detection in Structures*], Springer, 1st ed. (2009).

[2] Farrar, C. R. and Worden, K., "An introduction to structural health monitoring," *Philosophical Transactions of the Royal Society, Series A* **365**, 303–315 (2007).

[3] Junga, U. and Koh, B.-H., "Structural damage localization using wavelet-based silhouette statistics," *Journal of Sound and Vibration* **321**, 590–604 (2009).

[4] Chakraborty, D., Kovvali, N., Wei, J., Papandreou-Suppappola, A., Cochran, D., and Chattopadhyay, A., "Damage Classification Structural Health Monitoring in Bolted Structures Using Time-frequency Techniques," *Journal of Intelligent Material Systems and Structures, special issue on Structural Health Monitoring* **20**, 1289–1305 (July 2009).

[5] Cohn, D. A., Ghahramani, Z., and Jordan, M. I., "Active learning with statistical models," *Journal of Artifcial Intelligence Research* **4**, 129–145 (1995).

[6] Caruana, R., "Multitask learning," *Machine Learning* **28**, 41–75 (1997).

[7] Thrun, S. and Pratt, L., eds., [*Learning To Learn*], Kluwer Academic Publishers, Boston, MA (1998).

[8] Chakraborty, D., Kovvali, N., Zhang, J., Papandreou-Suppappola, A., and Chattopadhyay, A., "Adaptive learning for damage classification in structural health monitoring," in [*43rd Asilomar Conference on Signals, Systems and Computers*], (November 2009).

[9] Escobar, M. D. and West, M., "Bayesian density estimation and inference using mixtures," *Journal of the American Statistical Association* **90**, 577–588 (1995).

[10] Ishwaran, H. and James, L. F., "Gibbs sampling methods for stick-breaking priors," *Journal of the American Statistical Association* **96**, 161–173 (2001).

[11] Papandreou-Suppappola, A., ed., [*Applications in Time-Frequency Signal Processing*], CRC Press, Florida (2002).

[12] Haykin, S., [*Neural Networks and Learning Machines*], Pearson Education Inc., 3rd ed. (2009).

[13] MacKay, D. J., "Information-based objective functions for active data selection," *Neural Computation* **4**, 590–604 (July 1992).

[14] Hara, K. and Nakayama, K., "Selection of minimum training data for generalization and online training by multilayer neural networks," in [*IEEE International Conference on Neural Networks*], **1**, 436–441 (June 1996).

[15] Chen, W., Liu, G., Guo, J., and Guo, Y.-J., "A new method for sample selection in active learning," in [*International Conference on Machine Learning and Cybernetics*], **4** (2009).

[16] Seo, S., Wallat, M., Graepel, T., and Obermayer, K., "Gaussian process regression: active data selection and test point rejection," in [*International Joint Conference on Neural Networks, Proceedings of the IEEE-INNS-ENNS*], **3**, 241 – 246 (2000).

[17] Mallat, S. G. and Zhang, Z., "Matching pursuits with time-frequency dictionaries," *IEEE Trans. on Signal Processing* **41**, 3397–3415 (December 1993).

[18] Chakraborty, D., Soni, S., Wei, J., Kovvali, N., Papandreou-Suppappola, A., Cochran, D., and Chattopadhyay, A., "Physics based modeling for time-frequency damage classification," in [*Proc. of SPIE, Smart Structures and Materials & Non-destructive Evaluation and Health Monitoring,*], **6926**, 69260M1–12 (2008).

[19] Chakraborty, D., Zhou, W., Simon, D., Kovvali, N., Papandreou-Suppappola, A., Cochran, D., and Chattopadhyay, A., "Time-frequency methods for structural health monitoring," in [*Sensor, Signal and Information Processing workshop*], (May 2008).

[20] Channels, L., Chakraborty, D., Butrym, B., Kovvali, N., Spicer, J., Papandreou-Suppappola, A., Afshari, M., Inman, D., and Chattopadhyay, A., "A comparative study of fatigue damage sensing in aluminum alloys using electrical impedance and laser ultrasonic methods," in [*Proc. of SPIE, Smart Structures and Materials & Non-destructive Evaluation and Health Monitoring,*], **7295**, 72950Q–1 – 10 (2009).

[21] Zhou, W., Chakraborty, D., Kowali, N., Papandreou-Suppappola, A., Cochran, D., and Chattopadhyay, A., "Damage classification for structural health monitoring using time-frequency feature extraction and continuous hidden Markov models," in [*Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers ACSSC 2007*], 848–852 (November 2007).

[22] Zhou, W., Kovvali, N., Papandreou-Suppappola, A., Peralta, P., and Chattopadhyay, A., "Progressive damage estimation using sequential Monte Carlo techniques," in [*7th International Workshop on Structural Health Monitoring*], (2009).

[23] Kuipers, L. and Niederreiter, H., [*Uniform Distribution of Sequences*], John Wiley & Sons Inc. (1974).

[24] Chazelle, B., [*The Discrepancy Method*], Cambridge University Press (2002).

[25] MacKay, D. J. C., [*Information Theory, Inference, and Learning Algorithms*], Cambridge University Press (2003).

[26] Boyd, S. and Vandenberghe, L., [*Convex Optimization*], Cambridge University Press, New York, NY (2004).

[27] Kullback, S. and Leibler, R., "On information and sufficiency," *The Annals of Mathematical Statistics* **22**(1), 79–86 (1951).

[28] Kullback, S., "The Kullback-Leibler distance," *The American Statistician* **41**, 340–341 (1987).

[29] Bhattacharyya, A., "On a measure of divergence between two statistical populations defined by their probability distributions," *Bulletin of the Calcutta Mathematical Society* **35**, 99–109 (1943).

[30] Kailath, T., "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Transactions on Communication Technology* **15**(1), 52–60 (1967).

[31] Yang, G. L., Cam, L., and Lucien, M., [*Asymptotics in Statistics: Some Basic Concepts*], Berlin: Springer. (2000).

[32] Weinzierl, S., "Introduction to monte carlo methods," (2000).

[33] Niederreiter, H., "Discrepancy and convex programming," *Annali di Matematica Pura ed Applicata* **93**, 89–97 (1972).

[34] Dobkin, D. P. and Eppstein, D., "Computing the discrepancy," in [*Proceedings of the ninth Annual Symposium on Computational Geometry*], 47–52 (1993).

[35] Dobkin, D. P., Eppstein, D., and Mitchell, D. P., "Computing the discrepancy with applications to supersampling patterns," *ACM Transactions on Graphics (TOG)* **15**, 354–376 (1996).

[36] Mohanty, S., Teale, R., Chattopadhyay, A., Peralta, P., and Willhauck, C., "Mixed Gaussian process and state-space approach for fatigue crack growth prediction," in [*International Workshop on Structural Heath Monitoring*], **2**, 1108–1115 (2007).