A Bayesian Hidden Markov Model-Based Approach for Anomaly Detection in Electronic Systems

Enkhjargal Dorj
Center for Advanced Life Cycle Engineering
University of Maryland
College Park, MD 20742
301-405-5323
enkhjargal@csms.edu.mn

Chaochao Chen
Center for Advanced Life Cycle Engineering
University of Maryland
College Park, MD 20742
301-405-5323
chaochao@calce.umd.edu

Michael Pecht, *Fellow, IEEE*
Center for Advanced Life Cycle Engineering
University of Maryland
College Park, MD 20742
301-405-5323
pecht@calce.umd.edu

Abstract—**Early detection of anomalies in any system or component prevents impending failures and enhances performance and availability. The complex architecture of electronics, the interdependency of component functionalities, and the miniaturization of most electronic systems make it difficult to detect and analyze anomalous behaviors. A Hidden Markov Model-based classification technique determines unobservable hidden behaviors of complex and remotely inaccessible electronic systems using observable signals. This paper presents a data-driven approach for anomaly detection in electronic systems based on a Bayesian Hidden Markov Model classification technique. The posterior parameters of the Hidden Markov Models are estimated using the conjugate prior method. An application of the developed Bayesian Hidden Markov Model-based anomaly detection approach is presented for detecting anomalous behavior in Insulated Gate Bipolar Transistors using experimental data. The detection results illustrate that the developed anomaly detection approach can help detect anomalous behaviors in electronic systems, which can help prevent system downtime and catastrophic failures.**

TABLE OF CONTENTS

1. INTRODUCTION 1
2. BAYESIAN HIDDEN MARKOV MODEL...........2
3. ANOMALY DETECTION APPROACH BASED ON BAYESIAN HIDDEN MARKOV MODEL CLASSIFICATION..4
4. APPLICATION OF THE ANOMALY DETECTION APPROACH...6
5. CONCLUSIONS..8
REFERENCES..9
BIOGRAPHY 10

1. INTRODUCTION

An anomaly in electronics can be defined as a deviation from normal behavior, and can be associated with parametric or non-parametric changes that evolve gradually over time. Early detection of anomalies in complex electronic systems prevents unexpected failures and enhances performance and availability [1]. The complex architecture of electronics, the interdependency of component functionalities, and the miniaturization of most electronic systems make it difficult to detect and analyze anomalous behaviors [2].

A Hidden Markov Model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. HMM-based classification technique offers an opportunity to determine unobservable hidden behaviors of complex and remotely inaccessible electronic systems using observable signals [3]. Dynamic nature of a system can be modeled as a Markov state model, in which measures for anomaly detection and prognostics can be formulated [4]. The Markov model presents system behavior better than any regression fit, especially for electronic systems in which failure that is not due to wear out mechanisms.

HMM-based health state monitoring techniques have been used for diagnostics and prognostics purpose in the past. For example, HMM-based health state monitoring techniques have been applied for health state detection of cutting tools in machining process [5][6][7][8], hydraulic pump [9][10] and gearbox of helicopter [3]. Chen et al. [11] developed a hybrid prognostic method of using high-order HMM and adaptive neuro-fuzzy system for carrier plate and bearing faults.

Smyth [12] pointed out that the Markov model transition parameters can be estimated from prior knowledge of the long-term system behavior and gross failure statistics. Baruah and Chinnam [6] highlighted that a Bayesian approach can be used to update the parameters of HMM-based prognostics model as new data becomes available. However, Bayesian HMM for anomaly detection has not yet been studied.

This paper presents a data-driven approach for anomaly detection of electronic systems based on a Bayesian HMM classification technique. The developed anomaly detection approach has two distinct phases: training and detection. In the training phase, HMMs are formulated that best describe the healthy and anomalous behaviors of systems. In the detection phase, the unknown observation sequences are then categorized as healthy or anomalous using the trained HMMs. This research uses Bayesian inference for estimation of the posterior HMM hyperparameters that are used to calculate the HMM model parameters. An application of the developed Bayesian HMM-based anomaly detection approach is presented for detecting anomalous behavior in Insulated Gate Bipolar Transistors (IGBTs) using experimental data. In Section 2, the mathematical background of Bayesian HMMs will be

978-1-4673-1813-6/13/$31.00 ©2013 IEEE

discussed. Section 3 will develop the Bayesian HMM classification-based anomaly detection approach, and Section 4 will present an example application of IGBT anomaly detection. Finally, the conclusions of this paper will be provided.

2. BAYESIAN HIDDEN MARKOV MODEL

HMM is a statistical model in which the system is assumed to be a Markov process with unobserved or hidden states. In an HMM, the state is not directly visible, but the output that is dependent on the state is observable. In defining HMMs, we will use similar notation as used by Rabiner [13]. We define a system at any time as one of a set of K hidden states, $S_1, S_2, \dots, S_K$. Depending on the set of probabilities associated with a state (called transition probabilities), the system may undergo the change of state at a discrete time. The time instants associated with state changes are represented as $t = 1,2, \dots, T$, and we denote the state at time t as q_t. The hidden state sequence can be denoted as $Q = \{q_1 q_2 \dots q_T\}$. The state transition probability matrix can be represented as A=$\{a_{ij}\}$, where $a_{ij} = P(q_t = S_j | q_{t-1} = S_i)$; $1 \le i, j \le K$; and $a_{ij} \ge 0$. The number of distinct observation symbols per state is M. Observation symbols correspond to the physical output of the system being modeled, which at any time can be one of a set of M observation symbols, $\{v_1, v_2, \dots, v_M\}$. We denote the observation at time t as O_t. The observation sequence can be denoted as $O = \{O_1 O_2 \dots O_T\}$. The probability of the observation symbol v_m having been produced from state j is, $B = \{b_j(v_m)\}$, where $b_j(v_m) = P(O_t = v_m | q_t = S_j)$, $1 \le j \le K$, and $1 \le m \le M$. The initial state distribution vector $\pi = \{\pi_i\}$, where $\pi_i = P(q_1 = S_i)$, $1 \le i \le K$. In compact form, the complete parameter set of the HMM can be represented as shown in Eq. (1):

$$\theta = \{A, B, \pi\} \tag{1}$$

The Bayesian HMM technique has been implemented in the past by some researchers. For example, MacKay [14] was the first researcher to apply variational methods to HMMs when the observations were discrete. Ji et al. [15] presented a Variational Bayes (VB) learning algorithm for continuous HMMs. McGrory and Titterington [16] applied variational methods in an HMM with Gaussian noise, which leads to an automatic choice of model complexity. Beal [17] applied the VB expectation maximization algorithm to HMMs and showed how model selection tasks, such as determining the dimensionality, cardinality, or number of variables, can be achieved by using VB approximations.

To define our Bayesian HMM, we specify Dirichlet distributions for the parameters $\theta = \{A, B, \pi\}$, with set of hyperparameters $\omega = \{\omega^{(A)}, \omega^{(B)}, \omega^{(\pi)}\}$. In Bayesian statistics, a Dirichlet distribution is considered as the conjugate prior of the categorical distribution. Conjugate prior distributions simplify the mathematical manipulations for performing Bayesian inference. Let us assume that in a model the data points follow a categorical distribution with unknown parameter vector θ and K number of categories. We treat the model parameter as a random variable and give it a prior distribution defined using a Dirichlet distribution with hyperparameter ω [17][18][19]:

$$P(\theta|\omega) \propto Dir(\omega) = \frac{1}{Z(\omega)} \prod_{i=1}^{K} \theta_i^{\omega_i - 1} \tag{2}$$

where, K is the number of categories, $\omega = \sum \omega_i$, $\omega_i > 0$, $Z(\omega) = \frac{\prod_{i=1}^{K} \Gamma(\omega_i)}{\Gamma(\omega)}$, and $\Gamma(\omega)$ is the gamma function.

In probability theory and statistics, a categorical distribution is a probability distribution that describes the result of a random event that can take on one of K possible outcomes, with the probability of each outcome separately specified. In our case, the observation sequence is categorical in nature. Hence, the likelihood of data (observation sequence) can be defined as follows [18][19]:

$$P(O|\theta) \sim Categorical(\theta) = \prod_{i=1}^{K} \theta_i^{C_i} \tag{3}$$

where $O = (O_1, \dots, O_T)$ the entire observation sequence and C_i is the count of number of times a random event of i_{th} category occurs in the observation sequence O.

According to Bayes rule,

$$P(\theta|O) \propto P(O|\theta) P(\theta|\omega) \tag{4}$$

Using Eqs. (2) and (3), the posterior parameters can be estimated by Eq. (5):

$$\begin{aligned} P(\theta|O) &\propto \left(\prod_{i=1}^{K} \theta_i^{C_i}\right) \left(\prod_{i=1}^{K} \theta_i^{\omega_i - 1}\right) \\ &= \left(\prod_{i=1}^{K} \theta_i^{C_i + \omega_i - 1}\right) \end{aligned} \tag{5}$$

Therefore, as per the conjugate prior method, if the prior is Dirichlet with parameter ω, then the posterior is Dirichlet with parameter $(C + \omega)$ [17][18][19]:

$$P(\theta|O) \propto Dir(C + \omega) \tag{6}$$

We set Dirichlet priors on our Bayesian HMM parameters A, B, and π with vector of hyperparameters $\omega^{(A)}, \omega^{(B)} and\ \omega^{(\pi)}$ as shown in Eqs. (7) - (9), where, $\omega_i^{(A)} = \{\omega_{i1}^{(A)} \dots \omega_{iK}^{(A)}\}$ is vector of hyperparameters for i_{th} row of parameter A, and $\omega_i^{(B)} = \{\omega_{i1}^{(B)} \dots \omega_{iM}^{(B)}\}$ is vector of hyperparameters for i_{th} row of parameter B.

$$p(A) = P(A|\omega^{(A)}) = \prod_{i=1}^{K} Dir(\omega_i^{(A)}) \tag{7}$$

$$p(B) = P(B|\omega^{(B)}) = \prod_{i=1}^{K} Dir(\omega_i^{(B)}) \quad (8)$$

$$p(\pi) = P(\pi|\omega^{(\pi)}) = Dir(\omega^{(\pi)}) \quad (9)$$

The expected count of i_{th} row of parameter A is represented by vector $C_i^{(A)} = \{c_{i1}^{(A)} \dots c_{iK}^{(A)}\}$, where K is the number of hidden states in the system, and $c_{ij}^{(A)}$ is the expected count of transitions from i_{th} state (S_i) to j_{th} state (S_j) given observation sequence O. The expected count of i_{th} row of parameter B is represented by vector $C_i^{(B)} = \{c_{i1}^{(B)} \dots c_{iM}^{(B)}\}$, where M is the number of distinct observation symbols that the system can produce, and $c_{im}^{(B)}$ is the expected count of the observation symbol v_m produced by i_{th} state given observation sequence O. The expected count of parameter π is represented by vector $C^{(\pi)} = \{c_1^{(\pi)} \dots c_K^{(\pi)}\}$, where K is the number of hidden states and $c_i^{(\pi)}$ is the expected count to be in i_{th} state at time $t = 1$ given observation sequence O. Therefore, the posterior of our Bayesian HMM parameters A, B, and π can be determined by the conjugate prior method as:

$$g(A) = P(A|O) = \prod_{i=1}^{K} Dir\left(C_{ij}^{(A)} + \omega_i^{(A)}\right) \quad (10)$$

$$g(B) = P(B|O) = \prod_{i=1}^{K} Dir\left(C_{im}^{(B)} + \omega_i^{(B)}\right) \quad (11)$$

$$g(\pi) = P(\pi|O) = Dir(C^{(\pi)} + \omega^{(\pi)}) \quad (12)$$

In order to estimate the posterior parameters, we need the vector of the expected counts $C^{(A)}, C^{(B)}$, and $C^{(\pi)}$, and vector of prior hyperparameters $\omega^{(A)}, \omega^{(B)}$, and $\omega^{(\pi)}$. The expected counts can be estimated by the forward-backward algorithm, along with the likelihood of the observation sequence (data). The forward-backward algorithm takes as input the mean of the parameters, whose distributions are defined by Eqs. (10) - (12). The mean of the parameters can be estimated by the digamma function. The digamma function is the standard function for estimating the geometric mean of Dirichlet distributions. The geometric mean values of the elements of the matrices of transition probability, emission probability, and initial probability, can be estimated as shown by Eqs. (13) - (15) [17], which results in sub-normalized probabilities.

$$a_{ij}^{mean} = exp\left[\psi\left(c_{ij}^{(A)} + \omega_{ij}^{(A)}\right) - \psi\left(\sum_{j=1}^{K}(c_{ij}^{(A)} + \omega_{ij}^{(A)})\right)\right] \quad (13)$$

$$b_{im}^{mean} = exp\left[\psi(c_{im}^{(B)} + \omega_{im}^{(B)}) - \psi\left(\sum_{m=1}^{M}(c_{im}^{(B)} + \omega_{im}^{(B)})\right)\right] \quad (14)$$

$$\pi_i^{mean} = exp\left[\psi(c_i^{(\pi)} + \omega_i^{(\pi)}) - \psi\left(\sum_{i=1}^{K}(c_i^{(\pi)} + \omega_i^{(\pi)})\right)\right] \quad (15)$$

where $\psi(x) = \frac{d}{dx} ln\Gamma(x)$ is the digamma function and $\Gamma(x)$ is the gamma function [17][18].

Likelihood of the observation sequence

The forward variable $\alpha_t(i)$ and backward variable $\beta_t(i)$ of the forward–backward algorithm are defined to evaluate the likelihood and the expected counts, $C^{(A)}, C^{(B)}$, and $C^{(\pi)}$. The forward variable $\alpha_t(i)$ is the probability of generating the partial observation sequence $O_1 O_2 \dots O_t$ at time t when system is in state S_i. The forward variable at time $t = 1$, $\alpha_1(i)$, is defined by Eq. (16) [13]:

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \le i \le K \quad (16)$$

The forward variable at time $t + 1$, $\alpha_{t+1}(j)$, can be calculated by multiplying the emission probability $b_j(O_{t+1})$ with the product of the forward variable and the corresponding state transition probabilities for all K states at time t, as shown by Eq. (17) [13].

$$\alpha_{t+1}(j) = b_j(o_{t+1}) \sum_{i=1}^{K} \alpha_t(i)\, a_{ij}, \quad 1 \le j \le K, \ 1 \le t \le T - 1 \quad (17)$$

The likelihood of the observation sequence, $P(O|\theta)$, is the sum of the terminal forward variables $\alpha_T(i)$, as defined in Eq. (18) [13].

$$P(O|\theta) = \sum_{i=1}^{K} \alpha_T(i) \quad (18)$$

The backward variable $\beta_t(i)$ is the probability of generating partial observation sequence $O_{t+1} O_{t+2} \dots O_T$ at time t and in state S_i, given the model. The backward variable at time T, $\beta_T(i)$, is defined as shown by Eq. (19) [13]:

$$\beta_T(i) = 1, \quad 1 \le i \le K \quad (19)$$

The backward variable at time t, $\beta_t(i)$, can be solved inductively as follows [13]:

$$\beta_t(i) = \sum_{i=1}^{K} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad 1 \le i \le K, \ t = T - 1, T - 2, \dots, 1 \quad (20)$$

Expected counts in the observation sequence

The posterior hyperparameters need to be re-estimated in each of the iterations of HMM training, since the Baum-Welch algorithm is an iterative learning algorithm. In order to update the posterior hyperparameters, variables $\xi_t(i,j)$ and $\gamma_t(i)$ need to be identified first. The variable $\xi_t(i,j)$ is the probability of being in state S_i at time t and in state S_j at time $t+1$, given the observation sequence $O_1 O_2 \dots O_T$ and defined as follows [13]:

$$\xi_t(i,j) = P\left(q_t = S_i, q_{t+1} = S_j \middle| O, \theta\right) \qquad (21)$$

$\xi_t(i,j)$ can be estimated using forward and backward variables, as shown by Eq. (22):

$$\xi_t(i,j) = \frac{\alpha_t(i)\, a_{ij} b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{K}\sum_{j=1}^{K} \alpha_t(i) \cdot a_{ij} b_j(O_{t+1})\beta_{t+1}(j)} \qquad (22)$$

The variable $\gamma_t(i)$ is the probability of being in state S_i at time t, given the observation sequence $O_1 O_2 \dots O_T$ and model θ, defined as [13]:

$$\gamma_t(i) = P(q_t = S_i | O, \theta) \qquad (23)$$

The probability $\gamma_t(i)$ can be expressed in terms of $\alpha_i(t)$ and $\beta_i(t)$ as [13]:

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{K} \alpha_t(i)\beta_t(i)} \qquad (24)$$

The expected count of transitions from i_{th} state (S_i) to j_{th} state (S_j), denoted by $c_{ij}^{(A)}$, can be estimated by:

$$c_{ij}^{(A)} = \sum_{t=1}^{T-1} \xi_t(i,j) \qquad (25)$$

The expected count of the observation symbol v_m produced by i_{th} state (S_i) given the observation sequence O, denoted as $c_{im}^{(B)}$, is defined as:

$$c_{im}^{(B)} = \sum_{t=1 \; s.t \; O_t = v_m}^{T} \gamma_t(i) \qquad (26)$$

The expected count to be in state S_i at time $t = 1$ is expressed as $c_i^{(\pi)}$:

$$c_i^{(\pi)} = \gamma_1(i) \qquad (27)$$

Once the counts have been estimated as above, distributions of the posterior parameters A, B, and π can be defined as shown in Eqs. (10)-(12).

Convergence criteria

The VB procedure is a practical implementation of Bayesian learning for the true posterior probabilities of the model parameters. Instead of computing the true posterior probabilities of the model directly, the VB approximates the true posterior to a variational posterior by maximizing a negative free energy. Each of the iterations either increases the negative free energy or leaves it unchanged, until it converges to a local maximum [15]. The negative free energy, F, is an important quantity to maximize the marginal likelihood and can be defined as shown in Eq. (28) [15][18][20]:

$$F(\theta) = \int g(\theta) \log p(O|\theta) d\theta - KL[g(\theta)||p(\theta)] \qquad (28)$$

where, the first term is the average likelihood of the data, and the second term $KL[g||p]$ is the Kullback–Leibler (KL) divergence between the approximating posterior g and the prior p, and given by Eq. (29) [18][20].

$$KL[g||p] = \int g(\theta) \log \frac{g(\theta)}{p(\theta)} \, d\theta \qquad (29)$$

KL is a positive quantity. When KL is greater than zero, F provides a lower bound on the model log-likelihood. When KL is equal to zero, F becomes equal to the model log-likelihood, and $g(\theta)$ is equal to the true posterior $p(\theta)$, and convergence is achieved.

3. Anomaly Detection Approach Based on Bayesian Hidden Markov Model Classification

This section develops the Bayesian HMM-based anomaly detection approach. The data pre-processing technique used to generate observation sequences, learning algorithm, and health state detection process are explained.

Data pre-processing

Data pre-processing in machine learning process helps to resolve issues such as noisy data, redundant data, and missing data values [21]. Data-preprocessing techniques such as instance selection and data discretization have been used in this approach. During the pre-processing phase, the first step is to average the data belonging to the same cycle to reduce the amount of data and the processing time. Averaging is one way to reduce the sheer size of a large data set, with very little loss of information [22]. Poritz and Richter [23] also used averaging to reduce the volume of computation. The next step in the pre-processing phase is data discretization. We discretized our data using the equal-width partitioning method that divides the range into *M* intervals of equal size. If *X* and *Y* are the lowest and highest values of the attribute, then the width of the intervals will be *W* = (*Y* - *X*)/*M*.

Bayesian HMM learning algorithm

Bayesian HMM learning algorithm is illustrated in Figure 1. Bayesian HMMs are created for each health state of the system with the model parameters treated as random variables of a Dirichlet distribution. In the initialization step, we initialize the hyperparameters of the HMM models by equal probabilities (uniform prior). The symmetric Dirichlet distribution is often used as a Dirichlet prior [17], and all elements of the parameter vector have a fixed uniform value. For example, if there are $K = 4$ hidden states, prior hyperparameters for A can be defined as $\omega^{(A)} = [0.25\ 0.25\ 0.25\ 0.25]$. The hyperparameters defined as such will be used as priors in the training iterations. Similarly, prior hyperparameters are defined for B and π. Random values for the Bayesian HMM parameters A, B, and π are generated from their respective Dirichlet distributions, as shown in Figure 1.

The initial counts are estimated by the product of random values of the parameters and the total length of the observation sequence. Bayesian inference of the variational posterior hyperparameters is computed using the conjugate prior method. The posterior hyperparameters are then used in the digamma function to estimate the mean of the Dirichlet distributions for the model parameters, i.e., the transition probability, emission probability, and initial probability. The mean of the Dirichlet distributions for the model parameters are then used in the forward-backward algorithm to estimate the expected counts and the log-likelihood of the observation sequence.

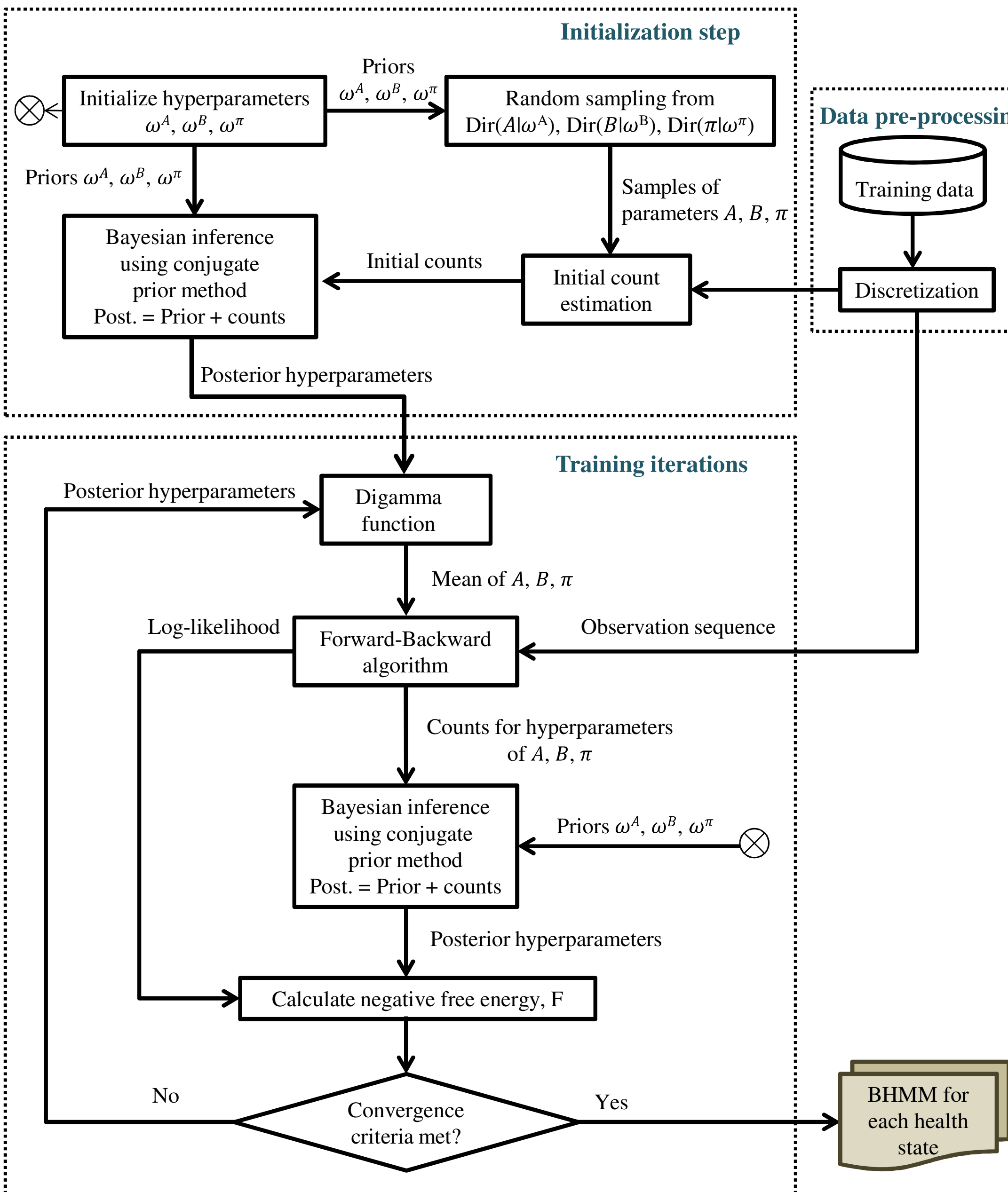


Figure 1- Flow chart of the Bayesian HMM training algorithm

The expected counts are used to estimate the posterior hyperparameters in the i_{th} iteration using the conjugate prior method (Eqs. (10) - (12)). The posterior hyperparameters and the log-likelihood of the data are used to estimate the negative free energy to determine whether the convergence criteria have been met. If the convergence criteria have not been met, then the next iteration is started, or else the training is stopped. The output of the training process is the posterior parameters of the Bayesian HMM.

The difference between Bayesian and regular HMM is that in regular HMM, the model parameters are estimated using the Baum-Welch algorithm, whereas in our Bayesian approach, the model parameters are treated as random variables of a Dirichlet distribution, and the posterior hyperparameters are estimated using a modified Baum-Welch algorithm. Further, regular HMM converges when the change in the likelihood for 2 consecutive iterations is less than the convergence coefficient, whereas in our Bayesian HMM approach, the convergence is established when the change in the negative free energy is negligibly small. The developed Bayesian HMM is based on VB, which approximates the true posterior to a variational posterior by maximizing a negative free energy.

Detection phase

In the detection phase, the discretization process used on the training data is also used for test data. The procedure for detecting the health state of an unknown observation sequence is illustrated in Figure 2. The likelihoods for healthy and anomalous HMMs are computed by the forward-backward algorithm with optimal posterior hyperparameters. The model with the highest likelihood is considered to be the model that represents the health condition of the unknown observation sequence.

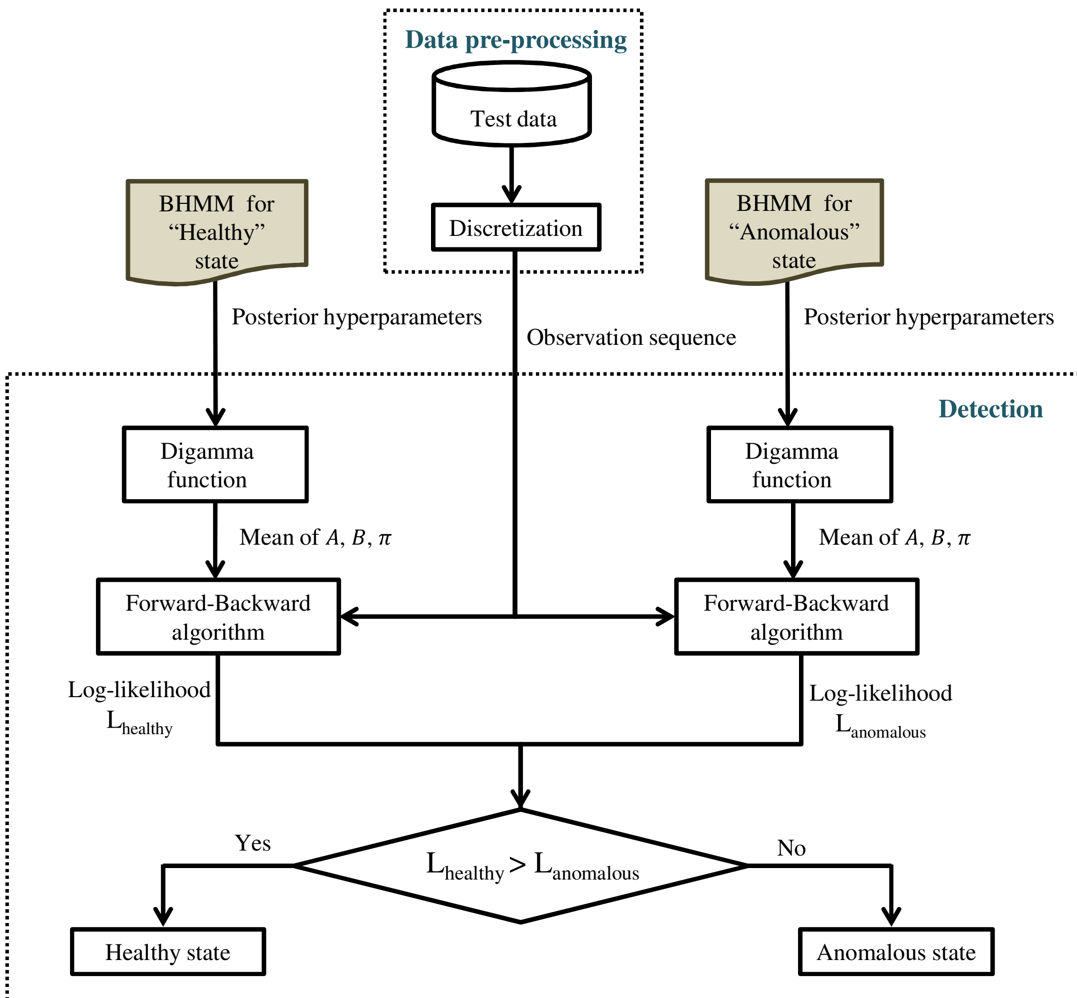


Figure 2- Flow chart of health state detection using Bayesian HMM

4. Application of the Anomaly Detection Approach

An application of the developed Bayesian HMM-based anomaly detection approach was demonstrated using experimental data for IGBTs. IGBTs are one of the most commonly used electronic products, and therefore were selected as a test-bed for validating the Bayesian HMM-based anomaly detection approach. Two health states, healthy, representing normal behavior, and anomalous, representing abnormal behavior, were monitored by our approach. Two HMMs representing the healthy and anomalous states needed to be trained. Table 1 presents the training samples used to train the Bayesian HMMs. The training samples have been obtained from three different operating conditions (1 kHz switching, 50% duty cycle, 100C swing; 5 kHz switching, 50% duty cycle, 100C swing; and 5 kHz, 60% Duty Cycle, 50C swing).

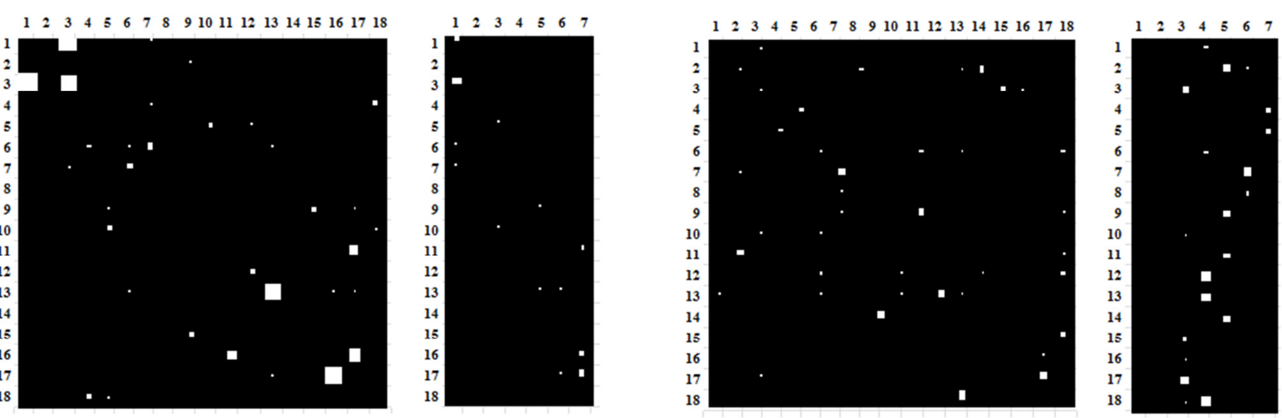
Table 1: Training samples

Training samples	Operating conditions	Size of training sequence	
Training sample 1 (8953x1) Number of cycle: 646	1 kHz switching, 50% duty cycle, 100C swing	Healthy (475x1)	Anomalous (171x1)
Training sample 2 (8187x1) Number of cycle: 482	5 kHz switching, 50% duty cycle, 100C swing	Healthy (121x1)	Anomalous (361x1)
Training sample 3 (15130x1) Number of cycle: 1178	5 kHz switching, 50% duty cycle, 100C swing	Healthy (103x1)	Anomalous (1075x1)
Training sample 4 (32867x1) Number of cycle: 4425	5 kHz, 60% Duty Cycle, 50C swing	Healthy (859x1)	Anomalous (3566x1)

The data pre-processing technique described in Section 3 was used to generate the training observation sequence. First, we averaged the data belonging to the same cycle to reduce the amount of data and the processing time. Then, equal-width partitioning was used to construct labeling schemes to supervise the training data for the healthy and anomalous models. Both the healthy and anomalous models need to have the same number of labels, M, in the observation sequence. The range of each observation sequence was divided into M intervals (partitions) of equal size. The number of labels M considered in the observation sequence is 7 based on the complexity and size of the available data.

Choosing the number of hidden states is an important problem for HMMs. Number of states within the HMM affects generalization of the model [6]. Too many hidden states can lead to over fitted model and poor performance. The hidden states can be chosen based on the knowledge of the system failure mechanisms [6]. However, the knowledge of failure mechanisms for complex and remotely inaccessible electronics such as IGBTs are often absent. The number of hidden states can also be chosen based on the knowledge of the observation sequences, i.e., the different types of observation sequences resulting from different combinations of available symbols/labels [17]. In this paper, we chose the number of hidden states K for the HMMs to be 18 based on the different observation sequences available. Further, we considered the same number of states for both the healthy and anomalous HMMs, since comparison of their likelihood is the basis of classification.

The Figure 3 presents the Hinton diagrams for the transition and emission matrices for the healthy and anomalous models. A Hinton diagram provides a qualitative display of the values in a data matrix (normally a weight matrix). From Figure 3a, it can be concluded that hidden states 8 and 14 are relatively inactive and have very low porbability of being transitioned into by other active states. Also, observation symbol 1 is mostly generated by the hidden state 3. Similar conclusions regarding the active states and observation symbols can be made from Figure 3b.



(a) (b)

Figure 3-Hinton diagram for posterior parameters A<18x18> and B<18x7> for a) healthy model (left), and b) anomalous model (right)

Figure 4 presents the learning curves used in the training process for optimizing the posterior parameters for both healthy and anomalous models. As is evident from Figure 4a, the negative free energy F value for the healthy HMM stabilize after around 50 iterations, whereas for anomalous model the negative free energy F value for the healthy HMM stabilize after around 25 iterations (Figure 4b).

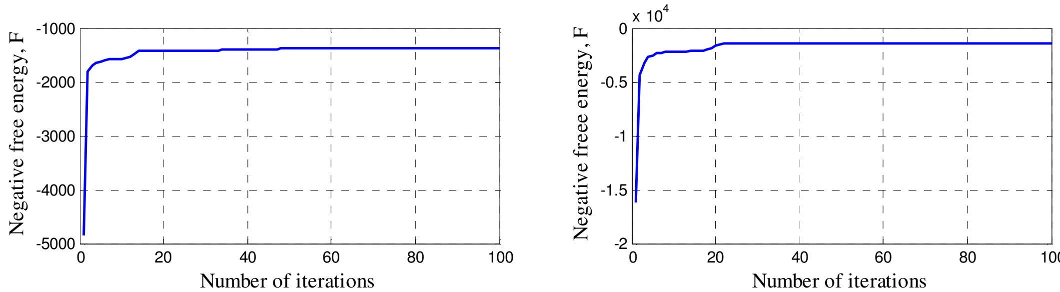


Figure 4a) Negative free energy for healthy model (left), and 4b) Negative free energy for anomalous model (right)

In the detection phase, 9 test sequences were used for health state detection using the healthy and anomalous models built in the training phase, as shown in Table 2. The test data 1 to 7 have been obtained from same operating conditions as training samples. Test data 8 and 9 have been obtained from different operating condition (1 kHz switching, 50% duty cycle, 50C swing), which has not been used for training. The model with highest likelihood is considered to be the model which represents the health condition of the test sequence. The detection accuracy of the Bayesian HMMs can be measured from Table 2 by comparing the last two columns, i.e., detected and real health state, which match for all but one of the data sets. As shown in Table 2, there is 88.9 % accuracy using the developed Bayesian approach. Further research is needed to improve the detection accuracy of the developed Bayesian HMM by using better priors for parameters and better health monitoring data.

Table 2: IGBT health state detection results using the developed Bayesian approach

Test cases	Operating conditions	Models to score test data	Log-Likelihood	Detected health state	Real health state
Test data 1 with 724 data points	1 kHz switching, 50% duty cycle, 100C swing	Healthy-BHMM	**-297.2707**	Healthy	Healthy
		Anomalous-BHMM	-495.5026		
Test data 2 with 166 data points	1 kHz switching, 50% duty cycle, 100C swing	Healthy-BHMM	-703.4825	Anomalous	Anomalous
		Anomalous-BHMM	**-87.5229**		
Test data 3 with 1101 data points	5 kHz switching, 50% duty cycle, 100C swing	Healthy-BHMM	**-205.4411**	Healthy	Healthy
		Anomalous-BHMM	-462.9955		
Test data 4 with 1758 data points	5 kHz switching, 50% duty cycle, 100C swing	Healthy-BHMM	**-802.2153**	Healthy	Healthy
		Anomalous-BHMM	-1306.7748		
Test data 5 with 73 data points	5 kHz switching, 50% duty cycle, 100C swing	Healthy-BHMM	-618.0505	Anomalous	Anomalous
		Anomalous-BHMM	**-43.2043**		
Test data 6 with 1727 data points	5 kHz switching, 60% duty cycle, 50C swing	Healthy-BHMM	-1799.4304	Anomalous	Healthy
		Anomalous-BHMM	**-1380.8795**		
Test data 7 with 3994 data points	5 kHz switching, 60% duty cycle, 50C swing	Healthy-BHMM	-106207.1074	Anomalous	Anomalous
		Anomalous-BHMM	**-231.0027**		
Test data 8 with 6972 data points	1 kHz switching, 50% duty cycle, 50C swing	Healthy-BHMM	**-290.6502**	Healthy	Healthy
		Anomalous-BHMM	-297.3739		
Test data 9 with 1277 data points	1 kHz switching, 50% duty cycle, 50C swing	Healthy-BHMM	-3485.4529	Anomalous	Anomalous
		Anomalous-BHMM	**-256.6022**		

5. CONCLUSIONS

In this paper, we present a data-driven approach for anomaly detection in electronic systems based on Bayesian HMM classification technique. The developed anomaly detection approach has two distinct phases: training and detection. In the training phase, Bayesian HMMs are formulated that best describes the healthy and anomalous behaviors of systems. The Bayesian HMM parameters are treated as random variables of Dirichlet distribution, the posterior hyperparameters of which are inferred using the conjugate prior method. The training process of the Bayesian HMMs is based on VB method, in which the true posterior is approximated to variational posterior by maximizing a negative free energy. In the detection phase, the likelihoods for healthy and anomalous Bayesian HMMs are computed by the forward-backward algorithm using the optimal posterior hyperparameters of the trained Bayesian HMMs. The model with highest likelihood is considered to be the model that represents the health state of the system corresponding to the unknown observation sequence.

An application of the developed Bayesian HMM-based anomaly detection approach was demonstrated using experimental data for IGBTs. Four data sets representing different operating conditions were used to train the healthy and anomalous models, while nine test observation sequences were used to test against the models. The detection result showed that the developed approach can help detect anomalous behaviors in electronic systems with 88.9% accuracy. Further research is needed to improve the detection accuracy of the developed Bayesian HMM by using better priors for parameters and better health monitoring data from field. The developed Bayesian approach also provides a framework for updating parameters when new data become available.

Acknowledgement

The authors would like to thank the more than 100 companies and organizations that support research activities at the Prognostics and Health Management Group within the Center for Advanced Life Cycle Engineering at the University of Maryland annually.

REFERENCES

[1] A. Ray, "Symbolic dynamic analysis of complex systems for anomaly detection," *Signal Processing*, vol. 84, no. 7, pp. 1115-1130, 2004.

[2] S. Kumar, N.M. Vichare, E. Dolev, and M. Pecht, "A health indicator method for degradation detection of electronic products," *Microelectronics Reliability*, vol. 52, no. 2, pp. 439-445, 2012.

[3] C. Bunks, D. McCarthy, and T. Al-Ani, "Condition-based maintenance of machines using hidden Markov models," *Mechanical Systems and Signal Processing*, vol. 14, no. 4, pp. 597-612, 2000.

[4] S. Kumar and M. Pecht, "Health monitoring of electronic products using symbolic time series analysis," *Proceedings of AAAI Fall Symposium on Artificial Intelligence for Prognostics,* Arlington, VA, 2007.

[5] H.M. Ertunc, K.A. Loparo, and H. Ocak, "Tool wear condition monitoring in drilling operations using hidden Markov models," *International Journal of Machine Tools and Manufacture*, vol. 41, no. 9, pp. 1363–1384, 2001.

[6] P. Baruah and R.B Chinnam, "HMMs for diagnostics and prognostics in machining processes," *International Journal of Production Research*, vol. 43, no. 6, pp. 1275-1293, 2005.

[7] F. Camci, "Process monitoring, diagnostics and prognostics using support vector machine and hidden Markov models," *PhD Thesis*, Wayne State University, 2005.

[8] F. Camci, and R.B. Chinnam, "Health-state estimation and prognostics in machining processes," *Transactions on Automation and Engineering*, vol. 7, no. 3, pp. 581-597, 2010.

[9] M. Dong, and D. He, "Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis," *European Journal of Operational Research*, vol. 178, no. 3, pp. 858-878, 2007.

[10] M. Dong, and Y. Peng, "Equipment PHM using non-stationary segmential hidden semi-Markov model," *Robotics and Computer-Integrated Manufacturing*, vol. 27, no. 3, pp. 581-590, 2011.

[11] C. Chen, B. Zhang, G. Vachtsevanos, and M. Orchard, "Machine condition prediction based on adaptive neuro–fuzzy and high-order particle filtering," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 9, pp. 4353-4364, 2011.

[12] P. Smyth, "Hidden Markov models and neural networks for fault detection in dynamic systems," *Proceedings of the IEEE-SP Workshop on Neural Networks for Signal Processing*, pp. 582–592, September 6-9, 1993.

[13] L.R. Rabiner, "A tutorial of hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257-286, 1989.

[14] D. MacKay, "Ensemble Learning for Hidden Markov Models," 1997, Available online at http://www.inference.phy.cam.ac.uk/mackay/ensemblePaper.pdf

[15] S. Ji, B. Krishnapuram, and L. Carin, "Variational Bayes for continuous hidden Markov models and its application to active learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 522-532, 2006.

[16] C. A. McGrory, and D. M. Titterington, "Variational Bayesian analysis for hidden Markov models," *Australian & New Zealand Journal of Statistics*, vol. 51, no. 2, pp. 227–244, 2009.

[17] M.J. Beal, "Variational algorithms for approximate Bayesian inference," *PhD Thesis*, Gatsby Computational Neuroscience Unit, University College London, 2003.

[18] C.M. Bishop, "Variational Inference," Chapter 10 In M.Jordan, J Kleinberg, and B. Scholkorf, Editors, *Pattern Recognition and Machine Learning*, Springer, New York, 2006.

[19] Z. Ghahramani, "An introduction to hidden Markov models and Bayesian networks," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 15, no. 1, pp. 9-42, 2001.

[20] W. Penny and S. Kiebel, and K. Friston, "Variational Bayes," Chapter 24 In K. Friston, J. Ashburner, S. Kiebel, T. Nichols, and W. Penny, Editors, *Statistical Parametric Mapping: The analysis of functional brain images*, Elsevier, London, 2006.

[21] S.B. Kotsiantis, D. Kanellopoulos, and P.E. Pintelas, "Data preprocessing for supervised learning," *International Journal of Computer Science*, vol. 1, no. 2, pp. 111-117, 2006.

[22] G. Foster, "Data reduction by averaging," *The Journal of the American Association of Variable Star Observers*, vol. 24, no. 2, pp. 117-121, 1996.

[23] A. Poritz and A. Richter, "On hidden Markov models in isolated word recognition," *Proceedings of the IEEE International Conference on Acoustics*, *Speech and Signal Processing*, Tokyo, Japan, 1986.

Biographies

Enkhjargal Dorj *received her B.S. and M.S. degrees in computer science from Mongolian University of Science and Technology (MUST), Ulaanbaatar, Mongolia. She is currently a PhD exchange student at Center for Advanced Life Cycle Engineering (CALCE), University of Maryland College Park. Her research interests focus on machine learning techniques for prognostics and health management (PHM), data mining, and enterprise resource planning (ERP).*

Chaochao Chen *is a member of the research staff at the Center for Advanced Life Cycle Engineering (CALCE). His research areas include fault diagnosis and failure prognosis, focusing on data-driven approaches such as machine learning and statistical methods, prediction uncertainty management, prognostics and health management (PHM) software implementation, verification and validation, fault tolerant control and their applications to robotics, electronics, battery and various mechanical systems. Prior to joining CALCE, Dr. Chen spent over three years at the University of Michigan and Georgia Institute of Technology as a research fellow, working in PHM areas in collaboration with multiple organizations in industry and the military. He has published over 20 technical papers including those in IEEE Transactions on Industrial Electronics, IEEE Transactions on Instrumentation and Measurement, Mechanical Systems and Signal Processing, Expert Systems with Applications, ASME Journal of Dynamic Systems, Measurement, and Control. He has served as a session chair and been invited to give talks for several reputed international conferences. He received his PhD in Mechanical Engineering from Kochi University of Technology, Japan, in 2007.*

Michael Pecht *(M'83–SM'90–F'92) received the M.S. and Ph.D. degrees in engineering mechanics from the University of Wisconsin, Madison.*

He is the Founder of Center for Advanced Life Cycle Engineering, University of Maryland, College Park, which is funded by over 150 of the world's leading electronics companies at more than U.S.$6 million/year. He is also a Chair Professor in mechanical engineering and a Professor in applied mathematics with the University of Maryland. He is the Chief Editor for Microelectronics Reliability. He has written more than twenty books on electronic product development, use, and supply chain management and over 500 technical articles. He consults for 22 major international electronics companies, providing expertise in strategic planning, design, test, prognostics, IP, and risk assessment of electronic products and systems.

Dr. Pecht is a Professional Engineer, an ASME fellow, a SAE fellow, and an IMAPS fellow. Prior to 2008, he was the recipient of European Micro and Nano-Reliability Award for outstanding contributions to reliability research, 3M Research Award for electronics packaging, and the IMAPS William D. Ashman Memorial Achievement Award for his contributions in electronics reliability analysis. In 2008, he was the recipient of the highest reliability honor, the IEEE Reliability Society's Lifetime Achievement Award. In 2010, he was the recipient of the IEEE Exceptional Technical Achievement Award. He served as a Chief Editor of the IEEE Transactions on Reliability for 8 years and on the Advisory Board of IEEE Spectrum. He is an Associate Editor for the IEEE Transactions on Components and Packaging Technology.