Online adaptive learning of Left-Right Continuous HMM for bearings condition assessment

F Cartella[1], T Liu[1,4], S Meganck[1,2], J Lemeire[1,4] and H Sahli[1,3]

[1] ETRO Department, Vrije Universiteit Brussel, Plainlaan 2, Brussels, Belgium
[2] CoMo Lab , Vrije Universiteit Brussel, Plainlaan 2, Brussels, Belgium
[3] Interuniversity Microelectronics Centre - IMEC, Leuven, Belgium
[4] FMI Dept., Institute for Broadband Technology - IBBT, Gaston Crommenlaan 8 (box 102), B-9050 Ghent, Belgium

E-mail: {francesco.cartella, tingting.liu, stijn.meganck, jan.lemeire, hichem.sahli}@vub.ac.be

Abstract. Standard Hidden Markov Models (HMMs) approaches used for condition assessment of bearings assume that all the possible system states are fixed and known a priori and that training data from all of the associated states are available. Moreover, the training procedure is performed offline, and only once at the beginning, with the available training set. These assumptions significantly impede component diagnosis applications when all of the possible states of the system are not known in advance or environmental factors or operative conditions change during the tool's usage. The method introduced in this paper overcomes the above limitations and proposes an approach to detect unknown degradation modalities using a Left-Right Continuous HMM with a variable state space. The proposed HMM is combined with Change Point Detection algorithms to (i) estimate, from historical observations, the initial number of the model's states, as well as to perform an initial guess of the parameters, and (ii) to adaptively recognize new states and, consequently, adjust the model parameters during monitoring. The approach has been tested using real monitoring data taken from the NASA benchmark repository. A comparative study with state of the art techniques shows improvements in terms of reduction of the training procedure iterations, and early detection of unknown states.

1. Introduction

This work relates to corrective maintenance [1, 2], which is based on condition monitoring and it is performed after a degradation modality emerges in a system, with the goal of preventing the system to fail. A fundamental requirement of this strategy is the detection of the abnormal behavior, that will lead to the system failure, at an early stage of its evolution in order to give enough time to perform the repairing operations. Moreover, we follow a data-driven approach [3] using sensor data provided by monitoring the system behavior.

Several data-driven models have been proposed in the literature. Among them Hidden Markov Models, have been successfully applied to condition monitoring of manufacturing machines [4, 5].

The common usage of Hidden Markov Models, especially for condition assessment applications, consists of dividing the degradation pattern of the historical data of the system under study into different levels, from normal to deeper degradation (i.e.: good, partially worn,

seriously worn, broken), and to train several HMMs each one corresponding to a segment of observation. The online condition assessment is then performed by selecting the HMM that, on new incoming data, gives the highest likelihood [6–10]. Another approach [11] consists of training a single HMM by using the sensor measurements corresponding to the entire lifetime of the system, by setting the number of states equal to the number of degradation stages contained in the training history. Using this approach, the model maps each degradation phase to a state of the HMM. The condition assessment is performed online on new incoming data by calculating the Viterbi path [12] which returns the most probable state sequence.

Such approaches assume that all possible system condition states are known a priori and that training data sets from the associated states are available. In addition, the training procedure has to be performed offline, and only once at the beginning with the available training set. For this reason, the trained models are able to recognize only the conditions contained in the training data. These assumptions significantly impede component diagnosis applications when

- It is difficult to identify and train all of the possible states of the system in advance
- Environmental factors or operative conditions change during the tool's usage.

Since industrial machines are constantly subjected to updates, changes of configurations or changes of working conditions, a need for models that are able to automatically adapt to new situations, by dynamically learning the structure and the parameters, is recently arisen among the industrial and research community.

Another aspect to consider is that the initial model parameters are usually set at random as initial guess of the training procedure that iteratively adjust them in order to maximize the likelihood function on the training data. However, the random initialization may lie far from the optimal point of maximum likelihood, causing a slow converge of the training algorithm or a sticking to a local maximum.

This work overcomes the above mentioned drawbacks, of most HMM-based condition assessment approaches, by combining Continuous Hidden Markov Models (CHMM) with a Change Point Detection (CPD) algorithm [13]. The CPD algorithm identifies the points of a time serie data at which the most significative changes, i.e. events, occur. The output of the CPD algorithm is a segmentation of the input signal, where each segment corresponds to a degradation state of the considered machine. CPD can be used in batch mode and in an on-line mode:

(i) During the training phase, the batch CPD algorithm is employed to segment the historical observation signals in order to map each detected segment to a state of the CHMM. This allows to find the initial number of states for the HMM. Then the initialization of the model parameters can be performed simply by counting the number of transitions from one state to the others, and estimating the mean and the variance of the observation for each segment/state.

(ii) During the online condition assessment, the online version of the CPD algorithm is employed to detect a new condition which was unknown during the training phase. In this case, a new state, representing the new condition, is added to the CHMM and when enough data coming from the new state are collected, the training procedure is executed again to adapt the model parameters to the new situation.

A similar approach has been introduced by Lee et al. [14], in which multivariate statistical process control is used to detect the occurrence of an unknown condition. However, a strong limitation is posed in the modeling of the observation signals, which are constrained to be gaussian distributed. This limitation is overcome by this paper, since the CPD algorithm is independent from the density family used to model the observation. Hence, an arbitrarily complex density can be used at this purpose.
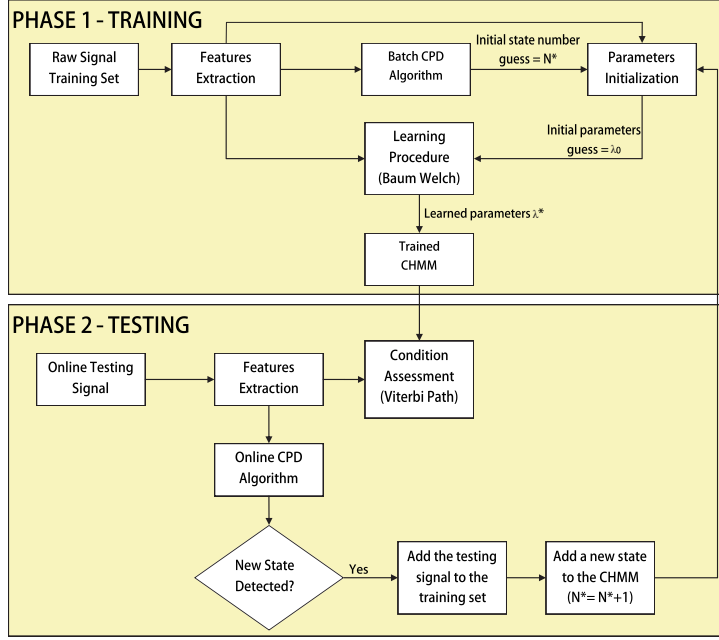
Figure 1. Framework of the proposed approach.

The proposed methodology is demonstrated on condition monitoring of bearings for which we can assume that the degradation pattern is monotonically increasing; hence left-right HMMs can be used. Validation using real data related to bearings taken from the NASA benchmark repository [15] is presented, along with comparison to state-of-art approaches.

The rest of this paper is organized as follows: Section 2 introduces the theoretical background of Continuous Hidden Markov Models and the Change Point Detection algorithm together with a detailed description of the proposed approach. In Section 3 experimental results are discussed. The conclusion and future research directions are given in Section 4.

2. Proposed methodology
This section gives a theoretical overview of the techniques used in this paper, together with the description of the proposed approach's framework, which is summarized in Figure 1.

2.1. Left-Right Continuous Hidden Markov Models
Left-Right Continuous Hidden Markov models (CHMMs) [11] are dynamic models in which there are two types of variables. A state variable describing the system state, which is never observed/measured directly (hidden), and an observation variable which is related to the hidden state.

A left-right CHMM $\lambda = (A, C, \mu, U, \pi)$ is characterized by the following:

(i) N, the number of states in the model, $S = \{S_1, \ldots, S_N\}$ the set of states, and s_t the state at time t.

(ii) The state transition probability distribution $A = \{a_{ij}\}$ where

$$a_{ij} = P(s_{t+1} = S_j | s_t = S_i), \qquad 1 \leq i, j \leq N. \tag{1}$$

(iii) $\mathbf{X}_t$, the continuous observation vector emitted from the state at time t.

(iv) The observation density, represented by a finite mixture:

$$b_j(\mathbf{X}) = \sum_{m=1}^{M} c_{jm} \aleph(\mathbf{X}, \boldsymbol{\mu}_{jm}, \mathbf{U}_{jm}), \qquad 1 \leq j \leq N \qquad (2)$$

where $\mathbf{X}$ is the vector being modeled, c_{jm} the mixture coefficient for the mth mixture in state j, and $\aleph$ is any log-concave or elliptically symmetric density (usually a Gaussian), with mean vector $\boldsymbol{\mu}_{jm}$ and covariance matrix $\mathbf{U}_{jm}$ for the mth mixture component in state j. The parameters c_{jm}, $\boldsymbol{\mu}_{jm}$ and $\mathbf{U}_{jm}$, are the coefficients of C, μ and U, respectively.

(v) The initial state distribution $\pi = \{\pi_i\}$ where

$$\pi_i = P(s_1 = S_i), \qquad 1 \leq i \leq N. \qquad (3)$$

It is proved that the probability density function of Equation 2 can be used to approximate, arbitrarily closely, any finite, continuous density function. Hence it can be applied to a wide range of problems.

The left-right topology has the property that the state index can either increase or stay the same, as the time goes on. In this case, the coefficients of the transition matrix are of the form

$$
\begin{aligned}
a_{ij} &= 0, \quad j < i & (4) \\
a_{ij} &= 0, \quad j > i + \Delta & (5) \\
a_{NN} &= 1 & (6) \\
a_{Ni} &= 0, \quad i < N. & (7)
\end{aligned}
$$

where Equation 4 avoids transition to previous states in the order, equation 5 avoids jumps of more than Δ states, while the state N is the last state, often called *absorbing* state. Moreover, since the model starts from state S_1, the initial state probabilities have the property

$$
\pi_i = \begin{cases} 0 & i \neq 1 \\ 1 & i = 1 \end{cases} \qquad (8)
$$

For CHMM's applied to condition monitoring, learning and inference procedures are used:

(i) (learning): Adjust the model parameters $\lambda = (A, C, \mu, U, \pi)$ in order to maximize the probability $P(\mathbf{X}|\lambda)$ i.e., for the considered CHMM model structure find the model parameters that better fit the historical sensorial measurements $\mathbf{X}$. This problem is solved through the Baum-Welch algorithm [16].

(ii) (decoding): Given a model $\lambda = (A, C, \mu, U, \pi)$ and a sequence of sensorial measurements $\mathbf{X} = \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_T$, find the hidden state sequence $S = s_1, s_2, \ldots, s_T$ that have most likely produced the observation sequence. The solution of this problem is given by the Viterbi algorithm [12].

2.2. Change Point Detection Algorithm

Change point detection approaches apply data mining techniques to identify the time points at which the changes, i.e. events, occur. This has been discussed in several applications: fraud detection [17], rare event discovery [18], event/trend change detection [13], and activity monitoring [19].

In this work we use the approach of [13], where two change point detection modes have been proposed: the batch (offline) and the incremental (online). Following the notation in [13], let $y(t), (t = 1, ..., n)$ be the time series to be segmented, with t the time variable. It is assumed that

the time series can be modeled mathematically, where each model is characterized by a set of parameters. Change-points detection is the problem of finding the function $\hat{f}_x(t, w_x)$ that best approximates the data, for each segment $x = [i, j]$. The quality of an approximation is measured by the loss function $\mathcal{L}(i, j) = Loss(y(t), \hat{f}_x(t, w_x))$ with $t \in x$, that has to be minimized. The loss function considered in this paper is the residual sum of squares.

Often, it is impossible to determine the nature of the approximation functions $\hat{f}_x(t, w_x)$'s from domain knowledge. As such, several types of basis functions can be considered, e.g. algebraic polynomials, wavelet, Fourier, etc. [20].

In the batch CPD, the entire data set is collected before the analysis begins. At every iteration, each segment is examined to see whether the time series can be split into two significantly different segments. The change-point in the first stage is the point j minimizing $\mathcal{L}(1, j) + \mathcal{L}(j + 1, n)$. Once this point has been found, the above procedure is repeated recursively on the resulting two segments, until a stopping criterion is reached. Formally, the algorithm should stop if in iterations l and $l + 1$

$$(\mathcal{L}_l - \mathcal{L}_{l+1})/\mathcal{L}_l < s \tag{9}$$

where s is a user-defined *stability threshold*.

Conversely, the online CPD is executed concurrently during data collection. Suppose that the last change-point was detected at time t_{l-1}. At time t_l the algorithm starts by collecting enough data to fit the regression model. Suppose at time t_j a new data point is collected. The candidate change-point is found by determining t_i such that $\mathcal{L}_{change} = \min_{l < i \leq j} \mathcal{L}(l, i) + \mathcal{L}(i + 1, j)$. The value $\mathcal{L}_{change}$ is then compared with $\mathcal{L}_{no\_change} = \mathcal{L}(l, j)$ and if

$$(\mathcal{L}_{no\_change} - \mathcal{L}_{change})/\mathcal{L}_{no\_change} > \delta, \tag{10}$$

then t_i is a change-point. Otherwise, the process should continue with the next point, i.e. t_{j+1}. δ being a user-defined threshold.

2.3. Framework Overview

As discussed in the introduction, the methodology is applied in two phases, the training and the online application. An extensive description of the methodology is given in the following for each of the two phases.

2.3.1. Phase 1: Finding the initial number of states and CHMM parameters initialization

During the first phase, we assume that enough historical sensor data have been collected from the monitoring system of the machine under study. Signal preprocessing is then performed to extract relevant information (features) correlated with the health condition of the system. In this paper, a vibration signal, collected through an accelerometer is used, and we consider as feature the mean value denoted as $y(t)$. This feature signal is then divided into N^* segments through the application of the batch version of the CPD algorithm. Each segment is mapped to a HMM state. The intuition of this approach comes from the assumption that there is correlation between observations and states. Thus, a variation in the condition of the system is related to a corresponding change in the behavior of the sensorial measurements.

Once the initial states are estimated, the initialization of the HMM parameters $\lambda_0 = (\pi_0, A_0, C, \boldsymbol{\mu}_0, \mathbf{U}_0)$ can be performed. Since left-right HMMs are considered, π_0 is initialized by using Equation 8. The transition matrix A_0 is initialized by using the following equations

$$a_{ij} = \frac{\#tr(i, j)}{\#s_i}, \qquad 1 \leq i \leq N^* - 1, \quad j = i + 1 \tag{11}$$

$$a_{ii} = 1 - \sum_{j \neq i} a_{ij}, \qquad 1 \leq i \leq N^* \tag{12}$$

being $\#tr(i,j)$ the number of transitions from segment i to j (which is equal to 1 for left-right HMM) and $\#s_i$ the number of data points in segment i. To ensure the validity of the left-right structure constraint, Equations 4, 6, 7 and 5 are also satisfied, with $\Delta = 1$ in Equation 5. This way, the transition matrix is ensured to satisfy the stochastic constraint, being $\forall i, \sum_{j=1}^{N^*} a_{ij} = 1$.

The observation density parameters μ_{im}, U_{im} and C_{im} (Equation 2) are estimated for each segment i as follows:

$$\forall m, \quad \mu_{im} \quad = \quad \overline{y}(t), \quad t \in i; \tag{13}$$

$$\forall m, \quad U_{im} \quad = \quad \mathrm{Var}(y(t)), \quad t \in i. \tag{14}$$

where $\overline{y}(t)$ and $\mathrm{Var}(y(t))$ are respectively the mean and the variance of the feature signal. The mixture coefficients C_{im} are initialized by using the uniform distribution, for each i and m: $c_{im} = 1/M$, M being the number of mixtures.

Together with the features extracted from the training histories, the estimated initial parameters λ_0 are used as inputs to the Baum-Welch procedure. The obtained optimal model parameters, λ^*, of the N^*-states left-right CHMM, are then used in the following *phase 2* for the online adaptive condition assessment.

2.3.2. Phase 2: Online adaptive condition assessment This phase is performed on new incoming data, referred to as testing signal in Figure 1. The incoming data are collected one sample a time, during the normal runtime of the machine under study. The above described features extraction (mean value estimation) procedure is performed each time a predefined number of samples are acquired.

For condition assessment, the feature signal is continuously input to the inference algorithm, using the N^*-states left-right CHMM model and its optimal parameters λ^* of *phase 1*, to calculate the most probable state sequence, given the observations up to that moment. The last state in the inferred sequence is then selected as the actual condition of the machine.

At the same time, the feature signal is also input to the online CPD algorithm to detect possible changes, which could correspond to state changes. Here, we assume that an unknown situation is experienced by the system when the detected segments (change points) exceed N^*, being the number of states of the optimal CHMM of the previous phase or iteration. In such case, when enough samples coming from the new condition are acquired, a new state is added to the CHMM, to obtain a new model with $N^* = N^* + 1$ states. The model parameters, λ^*, are also adapted to the new situation, by performing the training procedure as described in the above *phase 1*, using the testing signal together with the training histories.

3. Results
To assess the effectiveness of the proposed methodology, several experiments have been performed on condition monitoring data related to bearings taken from the NASA benchmark repository [15]. All the parameters concerning the CPD algorithms reported in this section have been carefully estimated through data analysis and selection of the optimal parameters.

3.1. Data description
The considered data [15] are monitoring measurements collected during several run-to-failure tests performed on bearings under constant load conditions on a special designed test rig. Four bearings (*bearing-1, ..., bearing-4*) were installed on a rotating shaft and the test has been performed by keeping a constant rotation speed of 2000 rpm, while a radial load of 6000 lbs has been added to the shaft and bearing by a spring mechanism.

All the bearings are forcedly lubricated and the failure of a bearing is detected by means of a magnetic plug installed in the oil feedback pipe, that collects debris from the oil as evidence

of bearing degradation. When the accumulated debris adhered to the magnetic plug exceeds a certain level, causing an electrical switch to close, a failure of at least one of the four bearings is assumed and the test will stop [21].

The data set consists of vibration signals collected by means of accelerometers. On each bearing housing, two accelerometers have been installed (one on the vertical Y axis and one on the horizontal X axis) for a total of 8 accelerometers.

Three run-to-failure tests have been performed on the test rig. One second snapshot data were collected at a sampling rate of 20 kHz (20480 samples for each snapshot) every 20 minutes for test-1 and every 10 minutes for test-2 and test-3.

At the end of the first run-to-failure experiment a crack was found near the shoulder of *bearing-3* while test-2 and test-3 ended up with an outer race defect in *bearing-1* and *bearing-3*, respectively.

3.2. Experimental results

Four histories of bearings that did not experience failure have been considered as training set. The raw data of the training set have been preprocessed to extract the mean value for each 20480 samples snapshot, resulting in a signal containing 2156 samples. Using these data, two segments have been detected by the batch CPD algorithm, parameterized with a polynomial fitting function of degree 1, and a stability threshold $s = 0.3$. Hence, a $N^* = 2$-sates left-right CHMM is considered.

To evaluate the effectiveness of the proposed initial parameters estimation approach on the training of the CHMM, 11 runs of the Baum-Welch algorithm have been made using as initial values the estimated λ_0 as described in Section 2.3.1, and 10 randomly set λ_0's. For these experiments, the number of mixtures has been set to three, which allows a trade-off between precision and computation time, and the maximum number of iterations has been set to 15. The training convergency speed, the final likelihood value, and the number of iterations, of the 11 runs are shown in Figure 2. As it can be seen, the learning procedure converges in only 3 iterations when the initial parameters are set using the proposed approach, compared to the randomly set parameters. Moreover, the final likelihood value is comparable with the best random initialization.
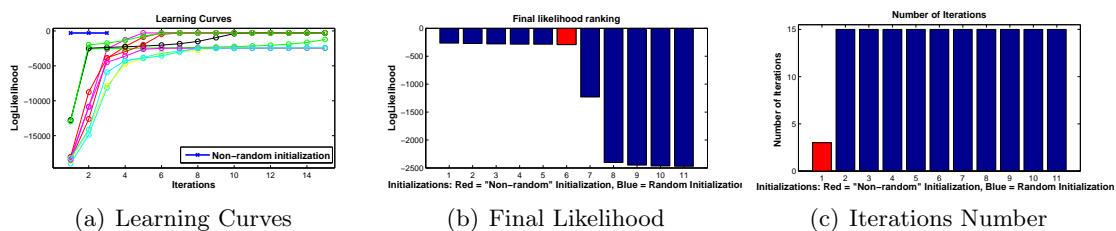

(a) Learning Curves   (b) Final Likelihood   (c) Iterations Number

Figure 2. Training Evaluation - proposed λ_0 estimation v.s. randomly set λ_0

The online condition assessment has been performed by collecting the testing history relative to a faulty bearing one point a time. However, as explained in Section 2.3.2, in order to perform the adaptive structure and parameter learning after the recognition of the unknown state, an amount of data coming from the faulty state are necessary. Since no experimental data have been collected after the breaking of the bearing, a simulated testing signal that resembles a bearing that breaks after 2158 time steps (see Figure 3) have been constructed, from which the mean values were estimated as features.

The online adaptive condition assessment of the above described *Phase 2* has been executed on the obtained testing signal, using the the above learned 2-states CHMM. For these experiments
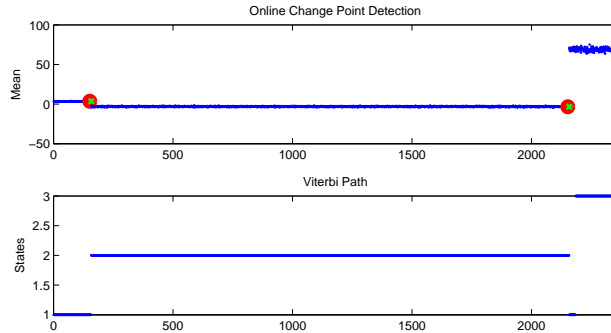
Figure 3. Proposed online adaptive condition assessment.
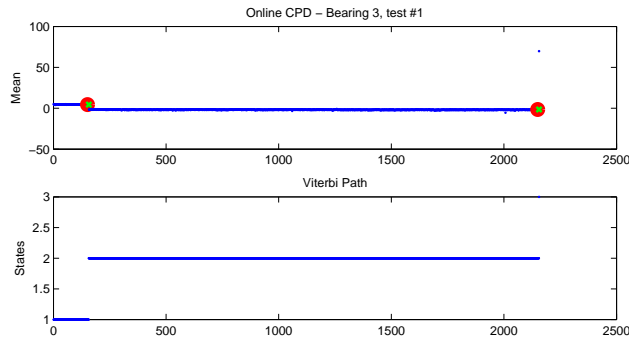

Figure 4. Faulty state detection for real bearing

a second degree polynomial function has been considered for the online CPD fitting function and a threshold $\delta = 0.6$. The obtained results are illustrated in Figure 3, depicting in the first row the change points detected by the online CPD (red dots), and in the second row the Viterbi path. As it can be seen, till sample 2159 the Viterbi path correctly classifies the real conditions of the bearing. Moreover, at that sample, the online CPD algorithm correctly detects a faulty state, which was previously unknown. 10 samples after, a new state is added to the 2-states CHMM. After learning the paraments of the new 3-states CHMM, classification of the third state is performed correctly until the end of the experiment.

Figure 3 also highlights the added value of the proposed adaptive procedure approach. Indeed, using the fixed structure CHMM, the Viterbi path is unable to give a correct estimation of the current condition at sample 2159, where a failure is experienced, and wrongly assigns the first state as the most probable state. Hence, the wrong classification would have been continued for the rest of the experiment if a fixed structure CHMM had been used in place of the proposed adaptive procedure.

The obtained 3-states CHMM has been tested on the real faulty bearing of the first run-to-failure experiment of the benchmark. As illustrated in Figure 4, the faulty state classification is correctly performed at the end of the experiment.

The above results have been compared to the ones using the approach of Lee et al. [14], in which the detection of a new state is performed by evaluating a statistically significant shift in the mean of the observation signal by using the Hotelling Multivariate Control Chart (referred as Hotelling MCC in the following). As in our approach we use a single feature as observation, a straightforward adaptation to the univariate case has been implemented.
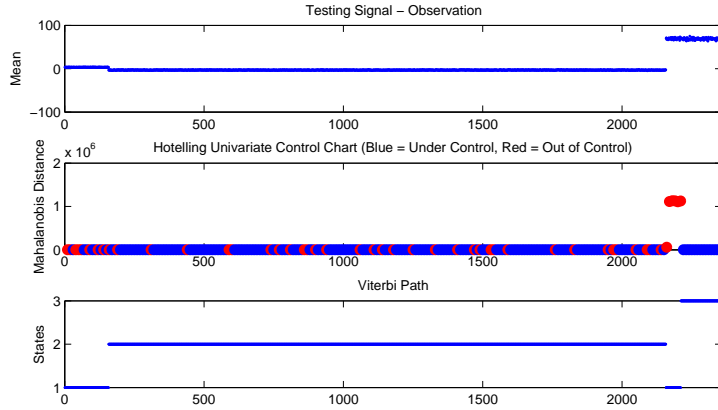
Figure 5. Online condition assessment using the Hotelling MCC.

In general, the Hotelling MCC considers as measure of similarity the Mahalanobis distance, denoted as $D^2(\overline{\mathbf{O}}, \boldsymbol{\mu}_i)$, between the empirical mean $\overline{\mathbf{O}}$ of the last b observations (set to 10 in this experiment) and the $\boldsymbol{\mu}_i$ parameter relative to the state i, which is the most probable state given by the Viterbi path:

$$D^2(\overline{\mathbf{O}}, \boldsymbol{\mu}_i) = b(\overline{\mathbf{O}} - \boldsymbol{\mu}_i)^T \mathbf{U}_i^{-1}(\overline{\mathbf{O}} - \boldsymbol{\mu}_i) \tag{15}$$

where $\mathbf{U}_i$ is the covariance matrix relative to the most probable state i. If this statistic exceeds a certain threshold (called Upper Control Limit, UCL) for more than R consecutive times, then a significant change is assumed in the process and a new state is added. R can be seen as a measure of the detection sensitivity of the algorithm, if R is increased, the detection algorithm may become more robust against false detections caused by the process randomness itself; on the other hand the algorithm may respond more slowly to the unknown state.

Concerning the UCL, since $\boldsymbol{\mu}$ and $\mathbf{U}$ are known, the D^2 statistic follows the χ^2 distribution with p degrees of freedom, where p is the number of multivariates, i.e., the number of features considered as observations. Thus, the UCL can be obtained as UCL $= \chi^2_{\alpha,p}$, where α is the risk level.

Figure 5 shows the complete execution of this procedure. The middle plot represents the Mahalanobis distance, with the points exceeding the UCL plotted in red. In this case R has been set to 5, and the detection of the failure modality happens at sample 2210, where a third state is added to the HMM. However, Figure 6 shows the limitation of this procedure represented by the modeling of the observations through a Gaussian distribution (and not with a mixture). The Mahalnobis distance is highly sensitive to small differences between the mean of the testing and the training observations, hence, it is applicable only when training and testing observations are very similar. This is demonstrated by the first part of Figure 6 where there are several consecutive points exceeding the UCL (denoted "Out Of Control") at the beginning because the mean of the training observations in this portion is slightly different from the testing set. If a smaller value of R had been considered, such method would have wrongly detected an unknown situation and a new state would have been added. However, at sample 2210 after the R consecutive points exceeding the UCL have been detected and the new state has been added to the HMM, the Mahalanobis distance is again reliable.

These results highlight that the proposed methodology (i) overcomes the limitations imposed by the Hotelling MCC since an arbitrarily complex density can be used to model the observations, and (ii) it is effective in detecting failure modalities at an early stage.
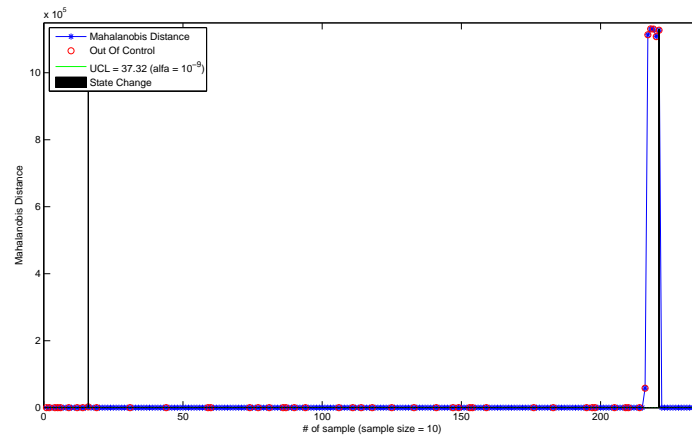
Figure 6. The Hotelling MCC for testing data.

4. Conclusion and future work

In this paper, a new approach to the online adaptive condition monitoring has been introduced. The methodology is based on continuous observation Hidden Markov Models that, combined with the Change Point Detection algorithm, are enhanced to deal with a variable state space, and are used for detection of unknown states and diagnosis. Moreover, a new approach for the estimation of the initial number of states of the CHMM has been provided, together with a method for parameters initialization. The obtained results illustrate that the proposed methodology (1) correctly estimates the initial number of states for the CHMM; (2) improves the execution speed of the training procedure in terms of iterations number reduction with a final likelihood comparable with the best random initialization; (3) detects an anomalous condition or an unknown state at an early stage; (4) changes the structure and the parameters of the CHMM to adapt the model to new situations.

In order to render the methodology more generally applicable, it is necessary to consider in future work multiple observations, as well as a method which is able to distinguish intermediate states and to recognize if the new detected state has already been encountered.

References

[1] Honkanen T 2004 *Modelling Industrial Maintenance Systems and the Effects of Automatic Condition Monitoring* Ph.D. thesis Helsinki University of Technology, Information and Computer Systems in Automation
[2] Wang H 2002 *European Journal of Operational Research* **139** 469–489
[3] Peng Y, Dong M and Zuo M J 2010 *The International Journal of Advanced Manufacturing Technology* **50** 297–313 URL http://www.springerlink.com/index/10.1007/s00170-009-2482-0
[4] Li Z, Wu Z, He Y and Fulei C 2005 *Mechanical Systems and Signal Processing* **19** 329–339
[5] Ertunc H, Loparo K and Hasan O 2001 *International Journal of Machine Tools & Manufacture* **41** 1363–1384
[6] McCarthy D and Al-Ani T 2000 *Mechanical Systems and Signal Processing* **14** 597–612
[7] Bjerkeseth M 2010 *Using Hidden Markov Models for fault diagnostics and prognostics in Condition Based Maintenance Systems* Ph.D. thesis University of Agder
[8] Zhu Y, Silva L C D and Ko C C 2002 *Pattern Recognition Letters* **23** 83–91
[9] Dimitrova N, Agnihotri L and Wei G 2000 *In European Signal Processing Conference*
[10] Liu T, Chen J and Dong G 2011 *Proc. of the 24th International Congress on Condition Monitoring and Diagnostics Engineering Management (COMADEM 2011)* pp 166–172
[11] Rabiner L R 1989 *Proceedings of the IEEE* **77** 257–286 ISSN 00189219 URL http://dx.doi.org/10.1109/5.18626
[12] Forney G D 1973 *Proceedings of the IEEE* **61** 268–278

[13] Guralnik V and Srivastava J 1999 *Proc. of KDD-99* pp 33–42

[14] Lee S, Li L and Ni J 2010 *Journal of Manufacturing Science and Engineering* **132** 021010 (pages 11) URL `http://link.aip.org/link/?MAE/132/021010/1`

[15] NSF-I/UCRC Center for intelligent maintenance systems. prognostic data repository: Bearing data set. In http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/, online in 2010

[16] Dempster A P, Laird N M and Rubin D B 1977 *Journal Of The Royal Statistical Society, Series B* **39** 1–38

[17] Burge P and Shaw-Taylor J 1997 *Proc. of AI Approaches to Fraud Detection and Risk Management* pp 9–13

[18] Yamanishi K, Takeuchi J, Williams G and Milne P 2000 *Proc. of KDD2000* (ACM Press) pp 250–254

[19] Fawcett T and Provost F 1999 *Proc. of KDD-99* pp 53–62

[20] Cherkassky V and Mulier F 1998 *Learning from Data* (New York, N.Y.: Wiley-Interscience)

[21] Qiu H, Lee J and Lin J 2006 *Journal of Sound and Vibration* **289** 1066–1090