

MACHINERY FAULTS DETECTION AND FORECASTING USING HIDDEN MARKOV MODELS

Paolo Calefati¹, Biagio Amico¹, Antonella Lacasella¹, Emanuel Muraca¹, Ming J. Zuo²

1 ITIA-CNR Institute of Industrial Technologies and Automation,, via delle Magnolie 4- 70026 Modugno
2 University of Alberta, Mechanical Department, Reliability Group

ABSTRACT

The present work describes an automatic procedure for diagnostics and prognostic issues, and its application to the evaluation of gearboxes residual lifetime. The Hidden Markov Models - HMM - technique has been used to create quasi-stationary and stationary models and to take advantages of the multiple sensor data acquisition architecture. At first, Markov models for diagnostics have been defined. The main advantage of the HMMs approach is that all vibration raw data measured by a multisensor architecture can be used without any pre-processing. An effort to adapt the HMMs technique to the prognostic issue has also been carried out. To create Markov Models suitable for prognostics, the Viterbi Algorithm has been used to define the best sequence of model states and to optimize residual useful lifetime computation. Finally, experimental results are discussed, which encourage further research efforts according to the proposed approach.

INTRODUCTION

One of the main methods for performing Condition Based Maintenance is based on vibration signals collected from a machine. The main objective is the detection of vibration characteristics which correspond to physical changes in the machine which indicate abnormal operation.

The primary challenge is to achieve a high degree of precision in classifying a machine's health given that its vibrational characteristics will vary with many factors not all corresponding to defective components. For example two identical new machines may have different vibrational characteristics due to differences in the manufacturing process.

It is important to differentiate between vibrational changes which are due to the machine component defects and those due to changing operating conditions.

In literature, several diagnostic techniques have been proposed in the past to detect the presence of fault in rotary machines. For such application, a Neural Network classifier seems to be an ideal candidate to correlate the input data to the presence of faults, thanks to its capability to learn complex and non linear mappings. Nevertheless, in most cases, an expert operator is needed to draw conclusions about the fault level by means of spectral analysis methods. Obviously, in order to reduce costs and simplify the diagnosis and prognostic stages, it would be desirable to make the fault detection and the estimation of residual lifetime fully automatic.

The Hidden Markov Models are suitable to perform detection and estimation operations for machine diagnostic and prognostic issues. In previous studies, these techniques have been applied to diagnose and forecast faults of mechanical components [1,2]. The motivation to use the HMM the Condition Based Maintenance comes from the success it provides in the area of speech processing. Today most commercial processing software tools for speech recognition, speaker identification and speaker verification are based on HMMs [3,4]. Such applications prove that HMMs are a successful instrument for speech processing. Furthermore, the processing method results robust to speaker variability and over populations of different speakers.

The major technical challenges in CBM which must be addressed by any processing methodology are summarized as follows.

In the CBM, machine vibration statistics are quasi-stationary and vary as a function of operating speed, torque, ambient or atmospheric conditions. CBM techniques must be able to distinguish between the normal changes and those due to defects. In the speech processing we have quasi-stationary

vibration statistics, varying as a function of the glottal source, the ambient or atmospheric conditions.

In the CBM, the machine character can be quite variable due to the differences in machining, part size, variations, fastener tightness, wear variations, replacement part variations, and aging. Machine monitoring techniques must be robust to these differences. In the Speech processing the voice character changes with the proper speaker voice.

Vibration features which are indicative of machine health can be hidden by vibration from other machine parts, a multiplicity of transmission paths, and ambient noise. Machine health monitoring techniques must be robust to multipath and must be effective in low signal -to-noise environments. Since there are many similarities between the speech processing and the processing for CBM applications, signal processing methodologies used in speech processing have good potential for successful application also in CBM.

In particular, one of the principal objectives of this research consists in evaluating the current condition of gearboxes, in order to estimate their residual useful lifetime. The effectiveness and reliability of the HMM to accomplish such a task is proven by experimental classification results on a gearbox system in different stages of the fault.

EXPERIMENTAL SETUP

In this section the experimental setup, used for the application of Hidden Markov Models to the problem of Condition Based Maintenance is presented. The data set, suitable to the prognostics issues, has been provided by the Department of Mechanical engineering, University of Alberta [5]. This data set consists of vibration measurements from a set of two accelerometers placed on the Machine Fault Simulator for the Gearbox, manufactured by Spectra Quest. At first a light fault was caused on one tooth of the gearbox 5 and after that an overload has been inducted on the machinery to speed up the fault's severity increasing. Data were collected during the whole gearbox working time from normal state to total fault state.

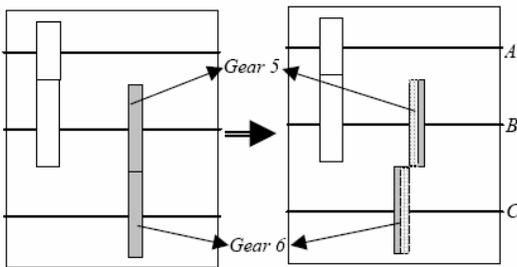


Fig 1- Gear 5 and gear 6 with slide off

The phenomena of the total fault are:

- Gear 5 and gear 6 have slide off
- Gear 5 and gear 6 have serious fault, especially gear 5
- Fault modes are teeth bent, teeth broken, and teeth missing.

The shaded area of the Fig.1 highlights the gear 5 shift. In this work, only the data relative to the slide off fault have been analyzed.

Fig.2 presents the Acquisition System Architecture used in this experiment, where the vertical accelerometer in position A and horizontal accelerometer in position B are highlighted.

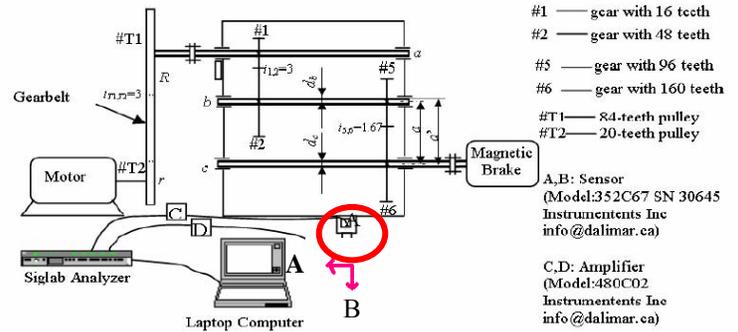


Fig 2 – The experimental setup

Two accelerometers have been used to collect simultaneous vibrational data. Such simple multi-sensor architecture is suitable for the required application. In fact we need of more sensors to be sure that our Hidden Markov Models are trained well to correctly recognize different situations of fault.

Two sensors is the simplest architecture to be sure to acquire the most significant data to detect the fault, because it is not possible to know a priori best accelerometer position .

HMM MODEL OF GEARBOX VIBRATION DATA

To monitor machines health two types of signal models can be adopted:

1. Deterministic models, where the specific properties of the signals are known through well characterized parameters like amplitude, frequency, etc., concerning the particular trend of wave.
2. Statistical models, characterized by statistical properties.

The machine fault phenomena are very changeable, so they can be characterized properly by means of stochastic models, using statistical parameters. In literature there are many different models to estimate the stochastic models parameters. One type of the stochastic signal model is the Hidden Markov Model. The elements of the Hidden Markov Model are:

- a. N , numbers of states $S = s_1 s_2, \dots, s_n$.
- b. M , numbers of the distinct observation symbols per state $Q = q_1 q_2 \dots q_m$.
- c. The state transition probability distribution $A = \{a_{ij}\}$,

where (a_{ij}) is the probability of moving from state s_i at time t , to state s_j at time $t+1$. It is an element of

$$A = P\left(q_{t+1} = s_j \mid q_t = s_i\right) \quad 1 \leq i, j \leq N .$$

- d. The observation symbol probability distribution in state j , $B = \{b_j(k)\}$, where

$$b_j(k) = P[v_{kat-t} | q_t = s_j], \quad 1 \leq j \leq N, \quad 1 \leq k \leq M$$

$b_j(k)$ is the probability staying in state j at time t .

e. The initial state distribution $\pi = \{\pi_i\}$, where,

$$\pi_i = P(q_i = s_i), \quad 1 \leq i \leq N.$$

A HMM is defined by $\lambda=(A, B, \Pi)$, giving the appropriate values of N, M, A, B and Π . The HMM can be used as a generator to give an observation sequence $\Theta=O_1 O_2 \dots O_T$ [3].

In this section the goal is to build HMM models which can be used to describe gearbox data. To build a correct Hidden Markov Model it is necessary to evaluate the following matrices [2]:

1. States Transition Matrix, A . The States transition matrix describes the probability to move from a stationary state to another one.

2. State Emission Matrix, B . The States Emission Matrix describes the statistics of the particular stationary model. Each element represents the probability distribution associated to each state.

To obtain these two components it is necessary to experimentally identify the frequency of occurrence of each defect and the average time spent operating at each condition. Usually such two components are not available, so a model to estimate the Transition Matrix and the Emission Matrix has to be set up.

Each state of the Markov chain associated to an HMM must have a state process model which specifies a state probability function. In the Hidden Markov Model literature the Gaussian distribution is often used although multi-modal distributions are also used by taking mixture Gaussian distribution [1, 2, 4, 6, 7, and 8]. Other choices are mixtures of autoregressive models and autoregressive moving average models [3, 7]. In the present work a simple multi-dimensional Gaussian mixture model has been used, because it is the best method for the adopted observation sequence.

The main diagonal of the covariance matrix, used to calculate the mixture Gaussian distribution, is built with the vibrational power RMS of each sensor data. The other elements of a covariance matrix are the RMS crosspower between the different sensors.

The acquired data demonstrate that the vibrational power RMS represent a good indicator for the health condition of the gearbox. As showed in the fig. 3, the RMS displays three different level of faults, thus, three classes have been modeled.

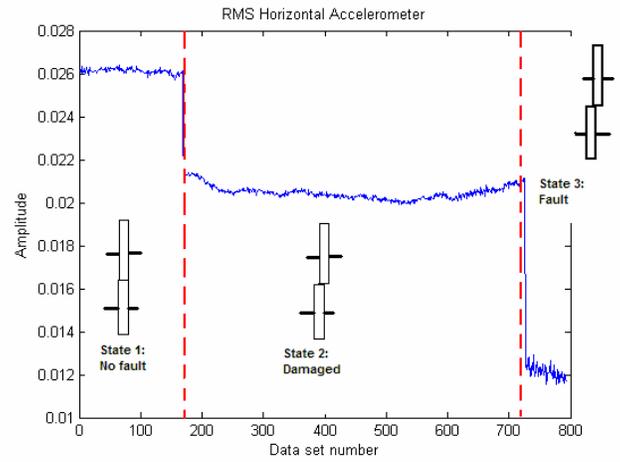


Fig 3 – Health condition of the gear box

The most important step is implementing the Hidden Markov Model for each state. So three HMMs have been implemented to be used as a dictionary. The aim is to obtain models that can be matched with new sequence of vibration data.

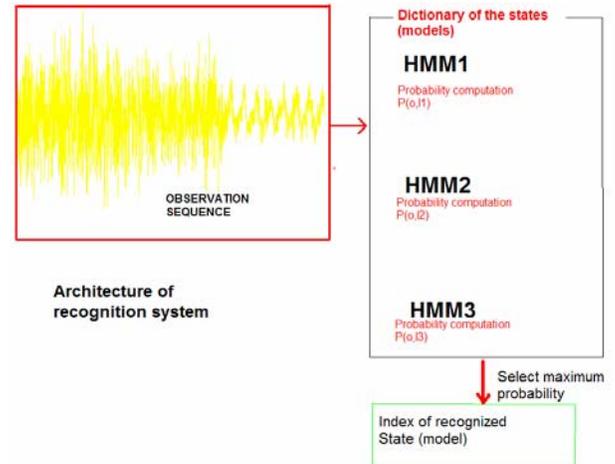


Fig 4 - Architecture of Recognition System

To create each model a training set of 60 arrays or patterns has been used. In Fig.5, the difference between the states can be seen. Typical vibration data for the state sequence from the healthy level to the faulty one are reported.

A different Gaussian is estimated for each one of the three cases. Each Gaussian is bidimensional (due to the fact that raw data from two sensors have been used), and is estimated using the first 1000 samples of each of the operating condition. The mean vectors μ_k and covariance matrixes Λ_k for each of the $K=1,2,3$ cases have been obtained, using the following formulas [3]:

$$\mu_k = \frac{1}{N} \sum_1^N y_k(n) \quad (1)$$

$$\Lambda_k = 1/N \sum_1^N [y_k(n) - \mu_k][y_k(n) - \mu_k]' \quad (2)$$

the Gaussian distribution formula is:

$$gm(x) = \sum_1^3 w_k g(\mu_k, \Lambda_k)(x), \quad K=1,2,3, \quad (3)$$

where $N = 1000$ samples and $y_k(n)$ is a 2D-vector of observations at time $n\Delta t$ for operating case k , w_k is the weight of the mixture and all weights are positive with the respect of $\sum_k^3 w_k = 1$.

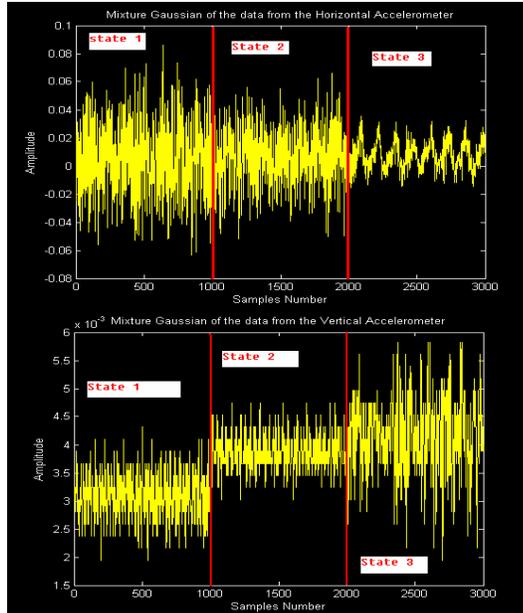


Fig 5 – Typical observation sequence

A different covariance matrix for each operating condition is used, and the likelihood of the analyzed data can be observed in the Fig 6, where the bidimensional Gaussian distribution for each case is built. The Gaussian Mixture let to evaluate the Transition matrix and Emission matrix belonging to each state.

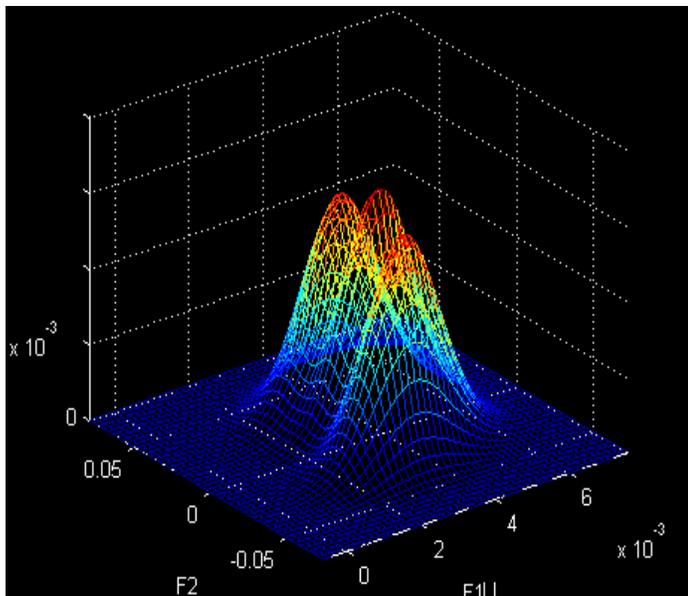


Fig 6 – The bidimensional Gaussian distribution

During the Markov Models training process, a model index has been assigned to the gaussians group belonging to each state. Each gaussian group is individuated by the arrays of acquired data at any detected component conditions. The training process gives to any arrays a weight w_k as bigger as that array is more similar to each other. In this way each gaussian distribution has its accurate space dislocation.

After founding the gaussian distribution it is very simple to found the array probability that such an array belongs to a determined model. These probabilities determine the emission matrix (probability to belong to a proper state) and the transition matrix (probability to belong to the other states) for each array. The three classes of Transition matrix and Emission matrix represent the Markov models concerning each health condition.

The accuracy of the HMMs has been tested in three classification experiments in which the number of samples of data from each sensor has been changed. The first test is lead with 1000 samples number of each data set from each accelerometer. Forty-one data sets for each state, different from the ones training the models, have been used. The following figure shows the results of the first test.

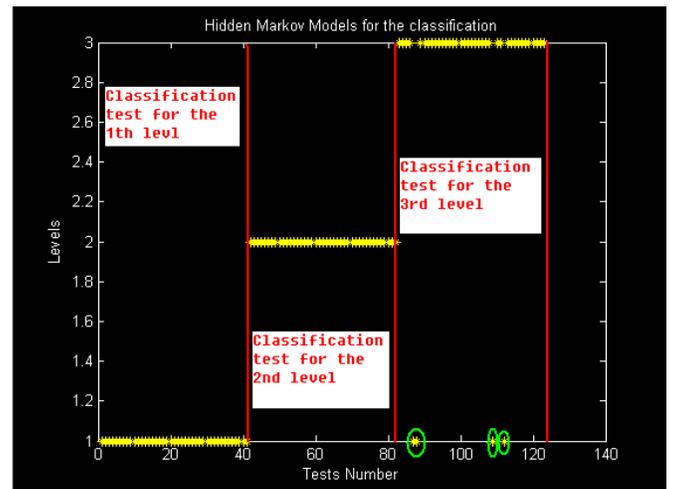


Fig 7 – Classification test, using the first 1000 samples of each array

As you can see in the figure above, the test has only three bad classification in the third levels, they are marked with green circles. The percentage of correct classification is 97%.

The second test is lead with 100 samples of each data set from each accelerometer and in this way you can see that the success of the correct classification is already decreased down to 95%. The third test is conducted decreasing the number of samples down to only 10. The results are coherent with the previous tests. The percentage of success of the correct classifications is decreased down to 89%.

The correct classifications decrease with the sample number of each observation sequence, that is due to the decreasing of the precision of the mean vector. In fact the mean vector plays an important role in training the function probability of the Gaussian distribution, also for calculating the covariance matrix.

Therefore, in this section, has been proved that is possible to apply the algorithms of the Hidden Markov Model to condition based maintenance and it can be a robust way to detect different conditions of gearboxes in laboratory environment.

THE HIDDEN MARKOV MODEL FOR PROGNOSTIC

An important problem in CBM is planning essential maintenance. Maintenance must be performed before the expiration of the remaining useful life of critical machine components. Thus the estimation of residual useful lifetime, known as prognostics, is an important and interesting problem.

In the previous section we assumed the importance of knowing the frequency of occurrence of each fault to understand the transition matrix. The transition matrix describes the probability to pass from one state to another one. The Hidden Markov Model can be trained by the use of time data series to obtain the optimal state sequence leading from the health state to the faulty one. It is important to know the exact sequence of the states to correctly predict the occurrence of a fault.

The optimal choice consists of selecting the state sequence (or path) that provides the maximum likelihood with respect to a given model. The sequence can be determined recursively via the Viterbi Algorithm [3]:

This algorithm makes use of two variables:

1. $\delta_n(i)$ is the highest likelihood of a single path among all the paths leading to state s_i at time n :

$$\delta_n(i) = \max_{q_1, q_2, \dots, q_{n-1}} p(q_1, q_2, \dots, q_n = s_i, x_1, x_2, \dots, x_n / \Theta)$$

2. a variable $\psi_n(i)$ which keeps track of the “best path” leading to state s_i at time n :

$$\Psi_n(i) = \arg \max_{q_1, q_2, \dots, q_{n-1}} p(q_1, \dots, q_n = s_i, x_1, \dots, x_n / \Theta)$$

where q_1, q_2, \dots, q_n is the observation sequence of the states, x_1, x_2, \dots, x_n are the values of observation sequence Θ .

The Trellis Diagram can be used to display the likelihood of calculations. Each column in the Trellis shows the possible states of machine health condition at a certain time n . Each state in a column is linked to the state of the adjacent columns by the transition likelihood given by the elements a_{ij} of the transition matrix A . At the bottom is the observation sequence $X = \{x_1, x_2, \dots, x_n\}$. $b_{i,k}$ is the likelihood of the observation $x_n = v_k$ in state $q_n = s_i$ at time n [3]. The following figure represents the Trellis diagram for only three states in which you can visualize likelihood calculations of HMMs.

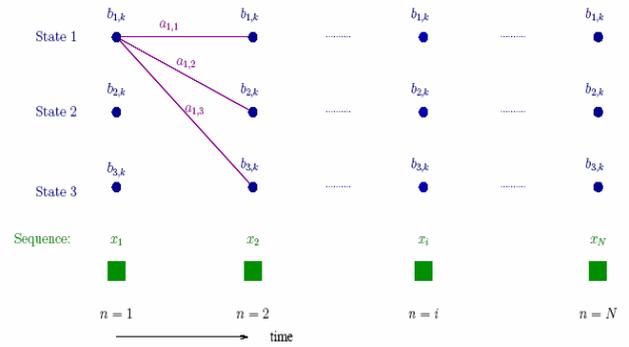


Fig 8 – The Trellis diagram

The idea of Viterbi Algorithm is to find the most probable path for each state in the Trellis Diagram. Every time only the most likely path leading to each state s_i ‘survives’.

The RMS time data series have been chosen to find the most probable state sequence.

Given these two curves a function proposing the Viterbi Algorithm for HMMs with the Gaussian emission has been implemented. After calculating the observation likelihood for these sequences, the δ and ψ vectors have been stored in a matrix in the format of the observation likelihood matrix. The main result is the assignment of a state for all point of our curves. The following figure shows the optimal state sequence for the RMS from the Vertical and Horizontal accelerometer, computed by the Viterbi Algorithm.

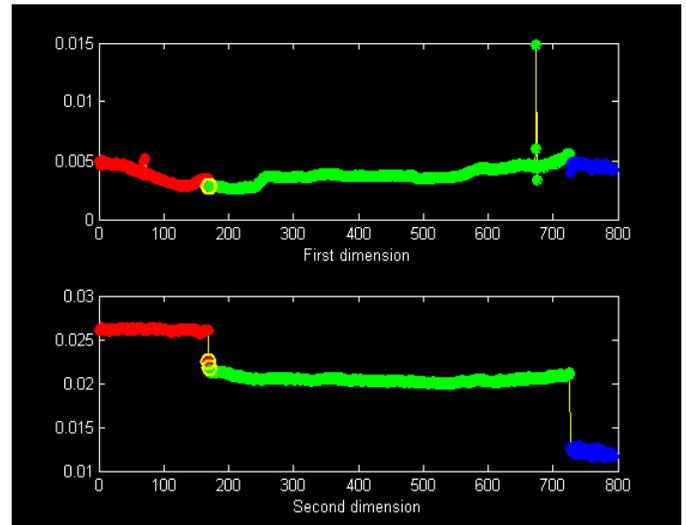


Fig 9 – Assignment of the optimal sequence of states to the RMS curves

The above figures show that the Viterbi Algorithm correctly determines the most probable path and makes a mistake in only two points, where assigns the state 1 instead of the state 2, as focused on the following figure.

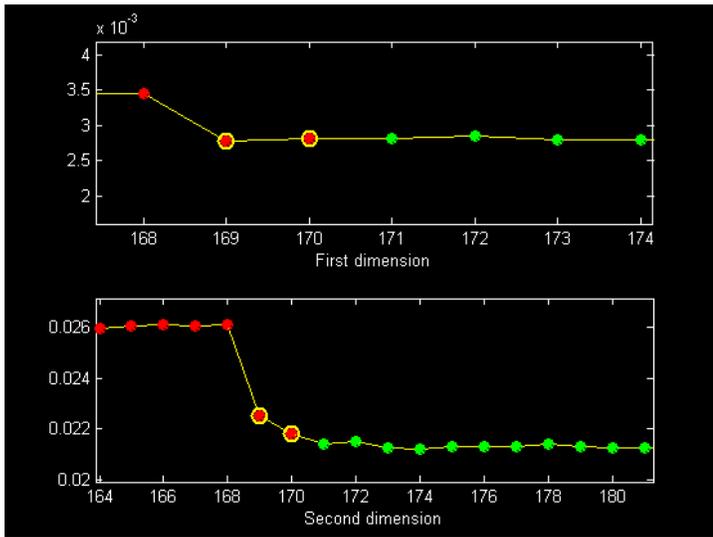


Fig 10 – Bad state assignment of two RMS points

The green state is the state two and the red one is the state 1. The points marked with the yellow circles are the misclassified states. The color blue in Fig.9 is the third state.

The Viterbi Algorithm gives the optimal state sequence or the path with the observation likelihood matrix. With the optimal state sequence and the observation likelihood matrix it is possible to build the Transition matrix in each point of the RMS observation [9].

For prognostic application, the most correct Markov Model is the left-to-right model, showed in the Fig.11

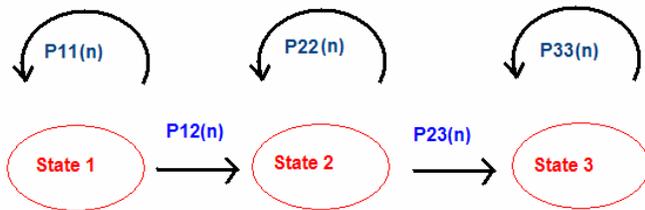


Fig 11 – Left-to-right Markov Model

The matrix associated to this model is the Transition Matrix:

$$\text{Transition Matrix} = \begin{bmatrix} P11(n) & P12(n) & 0 \\ 0 & P22(n) & P23(n) \\ 0 & 0 & P33(n) \end{bmatrix}$$

All the probabilities are function of the number of observation corresponding to the actual state

When the actual state corresponds to the first point of the RMS observation, the P11(n) is maximum and the P12(n) is minimum; the other probabilities in the matrix are constants. Moving forward the next points of observation, the P11(n) decreases and the P12(n) increases according to the sum P11(n)+P12(n)=1. Passing to the second state the probability

P11 and P12 become 0. In the second state you can repeat the same procedure with P22(n) and P23(n) at the same way. So n different transition matrixes can be built.

The Fig. 12 shows the function of the transition probabilities P12(n) and P23(n), varying the point of the RMS observation.

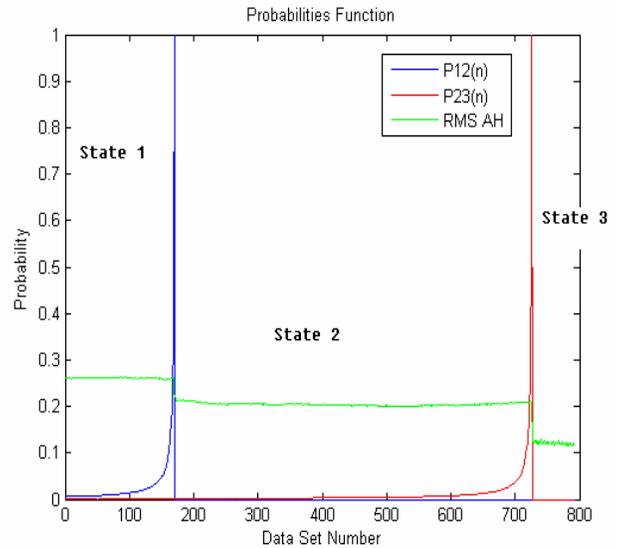


Fig 12 – Probability P12(n) and P23 (n) trends

The function built for prognostics, uses the detected data to obtain the observation likelihood matrix with the path. After that it matches these two parameters with those stored in a function of the RMS, and gives us the remaining useful lifetime. The correct path describes the actual state, the observation probability indicates in which time-point of the time series the system is, and the transition matrix gives information about the remaining useful lifetime concerning the previous state and point.

If the state N represents the state of zero remaining useful lifetime, then the mean time, t^* , to reach this state N from the actual n, is calculated as a function of the mean number of steps required to go from state n to state N.

$$t^* = \sum k p_k$$

where p_k is the probability of going from state n to state N in exactly k steps

Based on t^* , reasonable maintenance plan can be made.

CONCLUSIONS

In this work has been proved that the Hidden Markov Model is a good tool of analysis for the Condition Based Maintenance and it can be applied for diagnostic and prognostic issue.

Obtained results show that correct fault classification was obtained in 97% of cases, and that the estimation of remaining useful time is possible.

The HMMs present two main features particularly useful in monitoring machine health:

1. Computationally efficient methods (Continuous observation densities, Autoregressive HMMs, etc. [3]) exist for computing likelihoods using HMMs.

2. The HMMs can be used to build data-driven models of machines, able to identify health indicators of component defects and operating conditions.

The observation sequence for the HMM are completely general and can consist of many combinations of data features. So any defect detection algorithm can easily be integrated into an HMM formalism.

On going activities regard the application of the HMMs method to the detection and forecasting of bearing faults, using the data collected from the Machinery fault Simulator for bearing faults. As a matter of fact, by comparing the gearbox faults with the bearing ones, it may be noticed that there are many common points concerning the acquired vibration signal types and the opportunity to take advantage of the multiple sensor architecture of data acquisition. The experimental results encourage future work and show the validity of the proposed approach for the Condition Based Maintenance techniques and their applicability to many physical phenomena.

REFERENCES

[1] Carey Bunks and Dan McCarthy, Condition Based Maintenance of machines using Hidden Markov Models

Mechanical System and Signal Processing, **14**(4), 597-612, (2000).

[2] Mary Ostendorf and Gary D. Bernard, Hidden Markov Models for Monitoring machining tool-wear, Department of electrical engineer, University of Washington. Manufacturing Research & development Boeing Commercial Airplane Group 2000.

[3] L.R. Rabiner A tutorial on Hidden Markov Models and the speech processing, Proceedings of the IEEE **77**, 257-285, 1989

[4] Asmaa El Hannani, Dijana Petrovska-Delacrétaz, Raphael Blouet and Gérard Chollet, Segmental Score Fusion for Text-independent Speaker Verification. 2004.

[5] Hui Lin, Xianfeng Fan. The experiment results and Analysis of Gearbox Lifetime Prediction, Technical Report of Reliability group, Mechanical Department, University of Alberta, April,27,2004,

[6] L. A. Liporace, Maximum Likelihood estimation for multivariate observations of Markov sources, IEEE Transaction on Information Theory **IT-28**, 729-734, 1982

[7] A. B. Poritz, Linear Predictive hidden Markov models and the speech signal, Proceedings of ICASSP, Paris, France, May, 1291-1294, 1982.

[8] Reynolds, Quatieri and Dunn. Speaking Verification Using Adapted GMMs. 2000.

[9] Statistics toolbox, Matlab. Hidden Markov Models.