Optimal state selection and tuning parameters for a degradation model in bearings using Mel-Frequency Cepstral Coefficients and Hidden Markov Chains

Selección óptima de estados y parámetros de sintonización para un modelo de degradación de rodamientos utilizando Coeficientes Cepstrales en la Escala de frecuencias de Mel y Cadenas Ocultas de Markov

Mauricio Holguín^{1*} Álvaro Ángel Orozco¹ Germán A. Holguín¹ Mauricio Álvarez¹

Recibido 6 de octubre de 2014, aceptado 14 de enero de 2016

Received: October 6, 2014 Accepted: January 14, 2016

ABSTRACT

Preventive maintenance is a philosophy for assets management that aims to maximize operation through routine inspections with increasing frequency when no abnormalities are exhibit. This leads to an increase in the probability of failure due to the repetitive intervention and the inherent human error. Recently, forecasting research, or predictive research, have been addressed in order to obtain effective maintenance strategies and evaluate and manage the residual risk in equipment susceptible to degradation. Predictive research is related to the estimation of an active's Remaining Useful Life (RUL) by predicting its health state through the progression of its degradation. This article presents the development of an automated system that identifies types of faults in bearings, using Cepstral Coefficients on the Mel scale (MFCC) as the features set for description, and Hidden Markov Chains (HMC) with discrete observations as the classification method. Here we emphasizes on the optimal selection of the states in the HMM classifier using the ROC (Receiver Operating Characteristic) curves criteria. Features are discretized using clustering by k-means. Signals in this study are vibrations signals from the bearings in electrical machinery. The two databases used here are labeled with four different operation scenarios: normal, inner ring fault, outer ring fault, and rolling element fault. One of the databases allows for differentiation in severity levels for each scenario.

Keywords: Fault identification, feature extraction, Hidden Markov Chains, k-means clustering, Mel-Frequency Cepstral Coefficients, remaining useful life.

RESUMEN

El mantenimiento preventivo es una filosofía para la administración de activos que persigue el fin de maximizar la operación mediante rutinas de inspección, las cuales aumentan su frecuencia a medida que no se presenta un estado anormal, conllevando a que se incremente la probabilidad de falla debido a la mayor cantidad de intervenciones y por el error humano inherente. Recientemente la investigación de pronóstico ha sido utilizada con el objetivo de alcanzar estrategias efectivas de mantenimiento y con el fin de evaluar y administrar el riesgo residual en equipos que sufren degradación. El pronóstico se relaciona con la estimación de vida útil restante (RUL) de un activo al predecir su estado de salud durante la progresión de la degradación. Este artículo presenta el desarrollo de un sistema automático para identificar tipos de fallos en rodamientos, empleando Coeficientes Cepstrales en la escala de Mel (MFCC)

¹ Programa de Ingeniería Eléctrica. Universidad Tecnológica de Pereira. Complejo Educativo La Julita. Pereira, Colombia.
E-mail: mau.hol@utp.edu.co; aaog@utp.edu.co; gahol@utp.edu.co; malvarez@utp.edu.co

* Corresponding Author

como conjunto de características y Cadenas Ocultas de Markov (HMC) con observaciones discretas como clasificador. Se hace hincapié en la selección óptima de estados del clasificador HMM mediante el criterio de curvas ROC (Receiver Operating Characteristic). Las características son discretizadas empleando agrupamiento por k-medias. Las señales de estudio son datos de vibración provenientes de rodamientos. Se emplean dos bases de datos que presentan cuatro escenarios diferentes de operación, a saber: normal, falla en pista interna, falla en pista externa y falla en bola. Una de las bases permite diferenciar niveles de severidad para cada escenario de operación.

Palabras clave: Agrupamiento por k-medias, cadenas ocultas de Markov, coeficientes Cepstrales en la escala de Mel, extracción de características, identificación de fallos, vida útil restante.

INTRODUCTION

American manufacturers invested around USD 600 bi-million in maintenance of critical systems in the early eighties. These expenses doubled within just 20 years, and an alarming half of money was spent on non-effective maintenance [1]. However, the constant pressure on improving the reliability of assets remains intact [2]. So much that investments on preventive and predictive actions become globally 65.9% of maintenance resources in 2008 and 72.4% for the industrialized world [3]. What is worrying is that industrial energy consumption grows in global participation, with 42.6% in 2013 [4].

With the purpose of developing efficient maintenance strategies, challenges related to the predictive research have been faced to manage the residual risk of failing equipment [2].

Predictive research in this context should be understood as the estimation of Remaining Useful Life (RUL) for an asset by the prediction of the progression of a diagnosed anomaly [5]. Nonetheless, the diagnosis is basically a classification problem resorting to many methodologies addressed in the literature, as opposed to prognosis [6]. An efficient diagnosis, in order to make decisions about the structural health state, should deliver information regarding the location and severity following a detection procedure of events and statistical analysis of the observed characteristics [7].

Some normal strategies for fault diagnosis are based on visual observation of the spectrum with 2-D Fast Fourier Transform (FFT) and time-frequency contour graphs, which are unable to comply with the requirements of a prediction model [8]. The prediction models usually are trained with off-line

approaches, hence the trained model only respond properly to fault processes with equal features to the training data.

Consequently, these off-line models are not suitable to situations where features are changing. Moreover, the computational cost of training, prevent them from being used on-line for real time applications [9]. Therefore, the prediction models to date are theoretical and limited to small amounts of models and fault modes [2].

Another traditional strategy for RUL prediction is to use approaches based on the physical model of the mechanism, but they require specific knowledge of the system and generally do not reflect a general model for all the fault modes and for the entire life cycle of the mechanism [5, 10-11]. These models normally imply uncertainty from assumptions and simplifications due to the complexity and stochastic nature involved in the systems [11-13]. Furthermore, it is impractical to create a prediction model for the entire life cycle [14].

General interest has been focused on non-invasive techniques for fault detection and prediction, thus, basically implying signals-based techniques [15]. Parametric linear prediction techniques as autoregressive moving-average models (ARMA) usually works for short term predictions given the assumption of linearity of the process. Artificial Intelligence techniques has been used as well, as Artificial Neural Networks, however since they are black boxes and have slow convergence, they suffer from shortcomings such as difficulties of interpretation and structure.

Another prediction approaches, like Support Vector Machine (SVM), fuzzy logic, etc., have difficulties

to deal with stochastic problems because of the inherent randomness in most degradation processes [5, 13, 15]. Modal analysis has shown that modes can be insensitive to damage location and that changes in the modes forms can be unidentifiable due to noise and the limited number of recognizable modes forms [7].

Some applications where prediction is key part of the work, shows a model-based perspective applied in the early design and development of the asset. The model is simulated in order to improve reliability and availability. As an example of this case, in [16-17] the development of a model for a vehicle suspension is introduced. Diagnosis and RUL predictions are also included with projections of the system under different operation conditions.

Whether model-based or data-based, the general RUL inference methods rely on previous knowledge about how the system operates and the relative frequency of occurrence of the different kinds of defects. This perspective resembles the structure of a double stochastic process which must find, by probabilistic means, the total degradation state of a system and the probability of transition between states [18-19].

To a great extent, the aforementioned model is similar to those used in speech processing, since both have signals with quasi-stationary nature and in both cases there are differences in the parameters of similar characters from different systems. An example of such case is the variability in similar phonemes generated by the same vocal tract, or the variability in vibration signals from presumably similar machines under identical operation conditions.

Several speech processing systems use Hidden Markov Chains (HMC) since they allow for the analysis of a dynamic random process [18, 20, 21]. Given HMC hold the advantages of easy interpretation and the ability of performance in competitive learning environments, they are introduced as a method for classification of the residual life in a degradation process with the goal of characterize the health state of a mechanism [5, 6, 15].

Mel-Frequency Cepstral Coefficients (MFCC) provide signal information from both the time and the frequency domain, and they also enable

dynamic features extraction, whether they are linear or non-linear. These features are commonly used in vibration signals and speech processing [22, 23], and that i's the reason for us to use MFCC jointly with HMC in order to perform the bearings diagnosis.

MATERIALS AND METHODS

Data Base

We used two different data bases for the training and the validation of the proposed system. The first data base was compiled by the Bearing Data Center at Case Western Reserve University [24]. The one selected corresponds to a data acquisition of 12k samples per second and faults are induced through electric discharge of 0.1778, 0.3556 and 0.5334 millimeters in diameter to configure faults labeled according to three severity levels (low, medium, and high) in three elements of the bearings (inner ring, outer ring, and rolling element). More details on other parameters from this data base can be found in [24].

The second data base was developed by the Mechanical Vibrations Laboratory at Universidad Tecnológica de Pereira, with data acquisition at 20k samples per second. Faults are induced through mechanized action on the rolling element, the inner ring, and the outer ring.

Mel-Frequency Cepstral Coefficient (MFCC)

It is a parametric representation method based on the frequency response of the human hearing. MFCC have two filter types, linearly distributed for frequencies below 1 kHz, and logarithmically distributed for frequencies above 1 kHz [25]. This distribution is referred to as the Mel scale. The MFCC implementation process comprises seven steps detailed in Figure 1 [25].

- Segmentation: Dividing the signal into small frames with n samples each one. The frames usually overlap with the adjacent.

Figure 1. MFCC Implementation.

- Windowing: A Hamming window is usually used to adjust the frames and to integrate all the closest frequency lines.
- Pre-Emphasis: Filtering to emphasize the higher frequencies by increasing these frequencies' energy.
- FFT: Fast Fourier Transform using:

$$Y_m = \frac{1}{F} \sum_{n=0}^{F-1} x(n) w(n) e^{-j \frac{2\pi}{F} nm} \quad (1)$$

where F is the number of frames, w is the the Hamming Window function, and x is the input signal.

- Mel filters: The bank of filters distributed according to the Mel scale has a band pass triangular frequency response. The response magnitude is equal to unity at the Center frequency and decrease linearly to zero at center frequency of its two adjacent filters.
- DCT (Discrete cosine transform): The process to convert the Mel spectrum into time domain. The result of the conversion is called Mel Frequency Cepstrum Coefficients.
- Delta energy spectrum: given the quasi-stationary nature of the signals involved, the frames changes as they overlap in their transition. To represent this, 13 velocity deltas and 39 double deltas or accelerations are added.

Code Book

To use the MFCC feature set as discrete observations, a clustering technique such k -means is required [26]. The original procedure follows the guideline of defining k centroids in compliance to the existence of k clusters. The centroids are located so that they remain as far as possible from each other and every continuous observation point is associated to its nearest discrete centroid. Subsequently, several iterations are run in order to relocate the centroids and hence minimize the distance between continuous observations and centroids. This algorithm may have some variations directed to improve the computational efficiency [27]. The collection of centroids for different discretization processes are known as the Code Book.

Hidden Markov Chains (HMC)

A Markov Chain is any system that at certain instant of time can be considered to be in one of a finite

set of discrete states, denoted as $\{S_1, S_2, \dots, S_N\}$. At discrete uniformly distributed times, the system suffers state transitions according to a set of transition probabilities, for a time t at the current state q_t .

The chain is of first order if its transition probability only depends on the previous state, denoted as:

$$\begin{aligned} P\{q_t = S_j \mid q_{t-1} = S_j, q_{t-2} = S_k, \dots\} = \\ P\{q_t = S_j \mid q_{t-1} = S_i\} = a_{i,j} \end{aligned} \quad (2)$$

with $a_{i,j} > 0$ y $\sum a_{i,j} = 1$.

This model is known as an observable Markov Chain, since the process output is the same set of states at any time instant. On the other hand, a Hidden Markov Chain is the extension of the observable model, where the outputs are probabilistic functions of the state, and thus the model is an embedded double stochastic process which is not directly observable, but indirectly, through the set of output sequences [21, 28].

HMC has the following characteristics:

- N : number of states of the model
- M : number of symbols of the different observable states, denoted as $V = \{v_1, v_2, \dots, v_M\}$.
- $A = [a_{i,j}]$: Probability distribution of state transition.
- $B = [b_{i,j}(k)]$: Probability distribution of symbol observation at the state j , where $b_j(k) = P\{v_k \text{ en } t \mid q_t = S_j\}$.
- $\pi = [\pi_i]$: Initial state distribution where $\pi_i = P\{q_1 = S_i\}$.
- O : Observation sequence, where $O = \{O_1, O_2, \dots, O_T\}$, con $O_t \in V$.
- λ : Set of parameters for a HMC model, where $\lambda = \{A, B, \pi\}$.

With the assistance of HMC, three basic problems can be addressed, namely:

- The evaluation problem: given the observation sequence and a model, estimate efficiently $P\{O \mid \lambda\}$, *i.e.* the probability of observation sequence from the given model.
- Decodification problem: Given the observation sequence and a model, select the best state sequence that better explain the observations.
- Training problem: given the observations, adjust the model parameters to maximize $P\{O \mid \lambda\}$.

A description on how to address each of the aforementioned problems can be obtained from [21, 28].

RESULTS

Labelled Data bases

For training and validation, two data bases were used. One compiled by Bearing Data Center from Case Western Reserve University [24], and the other from the Mechanical Vibrations Laboratory of Universidad Tecnológica de Pereira.

For the Bearing Data Center database, the data was acquired at 12k samples per second, faults were induced in three different parts of the bearing (inner ring, outer ring and rolling element) with three levels of severity, one state of normal operation (base signal) and at four different velocities (1730, 1750, 1772 and 1797 RPMs.)

With the purpose of creating a labeled database, labels are assigned as follows: “Inner” for the inner ring faults, “Outer” for the outer ring faults, “Normal” for the base signal, “N1”, “N2” and “N3” for the 3 severity levels, N1 is the less severe and 3 is the most. The resulting database comprises ten folders: Normal, BallN1, BallN2, BallN3, InnerN1, InnerN2, InnerN3, OuterN1, OuterN2 y OuterN3.

For the database from the Mechanical Vibrations Laboratory, data were collected at 20k samples per second with induced faults in three parts of the bearing (inner ring, outer ring, and rolling element), one state of normal operation (base signal), and eighteen different operation velocities (200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 4000, 6000, 8000, 10000, 12000, 14000, 16000 y 18000 RPMs). The resulting basis comprises 4 folders: Normal, Ball, Inner, and Outer.

MFCC features and k-means extraction

To begin, the signal is divided in frames with lengths of 200 ms and with overlap 1/3. Afterwards, a Hamming window is applied in order to adjust the frames and integrate the closest frequency lines. Subsequent stages are the coefficients extraction, pre-emphasis, Mel filtering, DCT and delta energy spectrum. The preceding parameter

values are based on data reported by previous studies, as in [29-30].

The resulting vector represents a continuous scale for each possible characteristic. To train a HMC with discrete observations, a method to discretize features is required. In this context y observations (MFCC in this case) are vectors of d dimension (number of features) organized in $X = \{x_1, x_2, \dots, x_y\}$.

The k-means algorithm takes the observations and divided them into k sub-sets $S = S_1, S_2, \dots, S_k$. The algorithm iterates in two main steps [26]: In first place, observations that are close to each other are associated to means.

$$S'_i = \{x_p : \|x_p - m'_i\| \leq \|x_p - m'_j\|\} \quad (3)$$

Where $1 \leq j \leq k$. Then, an update is performed by estimating new measures for the centroids in each division:

$$m_i^{t+1} = \frac{1}{|S'_i|} \sum_{X_j \in S'_i} X_j \quad (4)$$

Finally, the algorithm converges when no significant changes occur in the actualization step.

HMC models training

To train the HMC models, the parameters $\lambda = \{A, B, \pi\}$ must be adjusted given the observations, in order to maximize $P\{O|\lambda\}$.

The Expectation Maximization (EM) algorithm is used, as in [31-32], where the probability of being on the state i at the time t and on the state j at the time $t + 1$, in order to estimate the parameters of the model, as:

$$\begin{aligned} \xi_t(i, j) &= P\{q_t = i, q_{t+1} = j | O, \lambda\} \\ &= \frac{P\{q_t = i, q_{t+1} = j | O, \lambda\}}{P\{O | \lambda\}} \end{aligned} \quad (5)$$

$\xi_t(i, j)$ can be also reformulated as a function of the two new variables $\alpha_t(j)$ y $\beta_t(i)$, as in:

$$\xi_t(i, j) = \frac{a_t(i) a_{i,j} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{i,j} b_j(O_{t+1}) \beta_{t+1}(j)} \quad (6)$$

where, the re-estimation of the parameters are:

$$\pi'_i = \sum_{i=1}^N \xi_i(i, j) \quad (7)$$

$$a'_{i,j} = \frac{\sum_{t=1}^{T-1} \alpha_t(i) a_{i,j} b_j(\mathbf{o}_{t+1}) \beta_{t+1}(j)}{\sum_{t=1}^{T-1} \sum_{j=1}^N \alpha_t(i) a_{i,j} b_j(\mathbf{o}_{t+1}) \beta_{t+1}(j)} \quad (8)$$

$$b'_j(k) = \frac{\sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^{T-1} \alpha_t(j) a_{i,j} b_j(\mathbf{o}_{t+1}) \beta_t = v_k}{\sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^N \alpha_t(j) a_{i,j} b_j(\mathbf{o}_{t+1}) \beta_t} \quad (9)$$

Iterations are performed until $P\{\mathbf{O}|\lambda'\}$, that is, while the new model is more probable to explain the observation sequence.

Model Evaluation

To evaluate the different HMC models that better represent the observations, we created Receiver Operating Characteristics (ROC) curves, which are broadly accepted as the *de facto* analysis and comparison method for diagnosis tests [33-35].

Sensitivity and Specificity are the measures for the diagnostic accuracy of a test. Sensitivity is defined as the rate of true positives and represents the proportion of observations that yield positive results on the test. On the other hand, specificity is defined as the rate of true negatives and represents the proportion of observations yielding negative results on the test.

An ROC curve represents a graphic of sensitivity against 1-specificity. The curve is obtained after performing a Montecarlo sampling on the solution set of the tuning parameters of the HMC model. When the model tuning distinguishes clearly the positive observations from the negative ones, the sensitivity will be 1 and the specificity 0 (i.e. $1 - \text{specificity} = 1$). When tuning is unable to make distinctions, sensitivity will be 0 and specificity 1 (i.e. $1 - \text{specificity} = 0$). The area under the ROC curve is used as measure of the diagnostic accuracy and as comparison criterion between the models [33, 35].

In the specific case of our HMC trained models, the number of states is the discriminant criteria, and the tuning parameters: the number of coefficients, the number of Mel filters, and the number of centroids

used in the discretization. The range of each variable includes the following values:

Coefficients: 1, 2, 3, 4, 8, 10, 12, 14, 16, 20.

Filters: 1, 2, 3, 4, 8, 10, 12, 14, 16, 20, 24.

Centroids: 2, 4, 6, 8, 10, 12, 14, 16, 20, 24.

Out of this total numbers of the parameters, it can be seen that a ROC is constructed with a total of 1100 points, each one corresponding to sensitivity and specificity values for a trained model with particular tuning parameters. This points are plotted in a normalized coordinate frame in the range from (0,0) to (1,1).

Figures 2 and 3 show ROC curves for HMC models with states 1, 2, 3, 4 and 5 over the data base from Case Western Reserve University Bearing Data Center with a 70 % of the data base used for training. Whereas Figure 2 shows the results when the data base reveals separation including severity levels, and Figure 3 shows the same revealed separation but with only faults state without severity.

Figure 4 show the ROC curves from HMC models with 1, 2, 3, 4 and 5 states for the Mechanical Vibrations Laboratory database with 70 % of the database used for training with fault states included but without severity levels since they are not available.

To provide a measure of diagnostic accuracy, ROC curves were smoothed using a moving average method and then, the area under each curve is

Figure 2. ROC curves for HMC trained with data from Bearing Data Center. Data base includes severity levels.

Figure 3. ROC Curves for HMC trained with data base from the Bearing Data Center. Data base includes fault states but severity is not available.

Figure 4. ROC curves for HMC trained with data from Mechanical Vibrations Laboratory of Universidad Tecnológica de Pereira.

estimated. Results are shown in Figure 5 for each of the set of curves in Figures 2, 3 and 4. However, the analysis has been expanded up to 7 states on each of the HMC models. Figure 5 show 3 curves denoting the areas under the ROC curve from the respective basis. Keep in mind that the curve labeled as “Bearing Data Center with severity” is the only one making separation that includes fault severity, whereas the other two (“Bearing Data Center without severity” and “Mechanical Vibrations Lab. - UTP”) only include separation by faults states.

Figure 5. Areas under the ROC curves for each HMC model trained and based on their respective data base.

To determine the optimal number of states for the HMC model, we look for a trade-off between the area under the ROC curve and the number of states used in HMC model. For the Bearing Data Center database case, Figure 5 shows area maximums for 3 and 6 states. Including severity levels, local maximums are achieved for 5 and 7 states. Accounting for computational complexity, it can be determined that 4 states reach values near the vicinity of the maximum area. For the case of the Lab. Mechanical Vibrations UTP database, Figure 5 shows a monotonically increasing area that is maximum at 7 states.

For the models whose data base does not include severity, the ROC area increases as the number of HMC states also increase. However when severity levels are included, the ROC area is prone to decrease after reaching its maximum. The larger the area under ROC curve, the greater the diagnostic accuracy and thus the discriminant power [36]. Despite some authors [37-38] suggestion on using the non-parametric Wilcoxon statistical test to analyze the differences under the areas, it can be shown that the concept of area under the ROC is closely related to Wilcoxon and is not affected by the probability distribution. This is especially convenient when working with probability distribution functions different to the normal.

From results just detailed above and with the purpose of obtaining a diagnostic accuracy with a lower computational cost, the number of states selected for the case is 4.

Model training

In order to determine the global reliability for the bearings degradation identification system, a 4-states HMC model was implemented using 24 coefficients, 24 Mel filters, and 32 centroids.

Each model was trained independently using 100 Montecarlo iterations, where the database was separated for training and testing with uniform probability distribution, and based on a heuristic cross-validation approach. The results for each of the three data bases are shown in Tables 1, 2 and 3. Results were divided into three training/testing separations as: 50%, 60%, and 70%.

Each row contains the results for a specific type of fault, including reliability and standard deviation. The last row consolidates the reliability of the total process for fault identification and/or severity fault levels.

Table 1. Training for the 4-states HMC using the Bearing Data Center data base without degradation levels.

	50%	60%	70%
Ball	0.9911 ± 0.0400	0.9736 ± 0.0715	0.9936 ± 0.0393
Inner	0.8750 ± 0.1727	0.8457 ± 0.1955	0.8082 ± 0.2047
Normal	1.0000 ± 0	1.0000 ± 0	1.0000 ± 0
Outer	0.7978 ± 0.2009	0.8293 ± 0.2100	0.8491 ± 0.2007
Total	0.9160 ± 0.0984	0.9121 ± 0.1009	0.9127 ± 0.1069

Table 2. Training for the 4-states HMC using the Bearing Data Center data base with degradation levels.

	50%	60%	70%
Ball N1	0.8583 ± 0.2766	0.9060 ± 0.2019	0.8850 ± 0.2231
Ball N2	0.8600 ± 0.3068	0.7760 ± 0.3599	0.7975 ± 0.3670
Ball N3	0.9617 ± 0.1250	0.9680 ± 0.1197	0.9675 ± 0.1160
Ball	0.8933	0.8833	0.8833
Inner N1	0.8900 ± 0.3145	0.9400 ± 0.2387	0.8100 ± 0.3943
Inner N2	0.9383 ± 0.1919	0.9360 ± 0.2144	0.9300 ± 0.2134
Inner N3	1.0000 ± 0	0.9980 ± 0.0200	1.0000 ± 0
Inner	0.9428	0.9580	0.9133
Normal	1.0000 ± 0	1.0000 ± 0	1.0000 ± 0
Outer N1	0.9867 ± 0.0698	0.9860 ± 0.0711	0.9800 ± 0.1161
Outer N2	1.0000 ± 0	1.0000 ± 0	1.0000 ± 0
Outer N3	0.8700 ± 0.3380	0.8800 ± 0.3266	0.9500 ± 0.2190
Outer	0.9522	0.9553	0.9767
Total	0.9365 ± 0.1406	0.9390 ± 0.1328	0.9320 ± 0.1447

Table 3. Training for the 4-states HMC using the Mechanical Vibrations Laboratory data base without degradation levels.

	50%	60%	70%
Ball	0.9720 ± 0.0550	0.9846 ± 0.0234	0.9813 ± 0.0232
Inner	0.9347 ± 0.0801	0.9329 ± 0.0697	0.9367 ± 0.0678
Normal	0.9999 ± 0.0006	1.0000 ± 0	1.0000 ± 0
Outer	0.9124 ± 0.0684	0.9047 ± 0.0739	0.9046 ± 0.0743
Total	0.9548 ± 0.0352	0.9556 ± 0.0360	0.9556 ± 0.0357

It should be highlighted here, that the Bearing Data Center database was considered twice. First, as a database without severity levels (Table 1) and second, as a database with severity levels, (Table 2).

CONCLUSIONS

The analysis of the areas under the ROC curves provided by Figures 2 to 5, leads to the conclusion that for both of the selected databases, four is the advisable number of states for which a HMC model for identifying degradation should be trained. From the training performed as described, it is noticeable on Tables 1 to 3, that variations on the percentage of the data base used for training do not alter significantly the total reliability for degradation identification. Tables report variations smaller than 0.007. For example, on Table 2, the difference between the lowest reliability total (when 70% of the data based is assigned to training), and the highest (when 60% of the data based is assigned to training) is only 0.9390 – 0.9320 = 0.007. Likewise, Table 1 and Table 3 yield difference of variations of 0.0039 y 0.0008 between each other. These variations provide evidence that HMC models have generalization power and are not prone to overtraining due to variations on the percentage of database separated for training.

The Bearing Data Center database achieve better results when the faults recognition includes severity levels, as seen in Table 2, the best result is achieved for a 60% of the database with a reliability of 0.9390 ± 0.1328. Conversely, when the same data base does not include the levels, the best result is reported on Table 1 for the 50% of the data base with a reliability of 0.9160 ± 0.0984. Table 3 reports the best result from the Mechanical Vibrations Laboratory database for the 70% of the data with reliability of 0.9556 ± 0.0357. The best performance

of the Bearing Data Center when including the degradation levels can be explained by the fact that each set of signals has greater variability when signals are not divided into severity levels.

Nonetheless, there exists the possibility of having to deal with the high number of models to be trained due to the division into degradation levels.

An important observation can be seen when comparing the two models using the Bearing Data Center database, the results without severity levels on Table 1 against the averages from severity levels of the same type of faults on Table 2. For both scenarios the Normal fault type is 100%. But the Ball fault type yields a better result when degradation is included, reporting a minimum average of 0.8833, versus 0.9736 reported without degradation. But the tendency changed for the Inner fault type and the Outer fault type, since the best results are given when degradation is not included: 0.8082 and 0.7978 respectively. Whereas the obtained minimum averages are 0.9133 for Inner and 0.9522 for Outer when degradation is included. In accordance with these comparison, the separation of classes and intra-classes seems affected by whether severity levels are included or not.

MFCC proved to be an efficient tool that allows obtaining either linear or nonlinear dynamics characteristics including time and frequency information.

A final remark is how clearly recognizable is the normal operation of the equipment. It can be easily separated from the faults types induced, as it can be seen on Tables 1, 2, and 3. This characteristic is encouraging to continue with the study of classification and identification of degradation and the severity levels on this kind of assets. On Tables 1, 2, and 3, which represent data bases without severity levels, the best ratio of results was reported for the fault induced on the rolling elements (referred to as "Ball" on the Tables). On the other hand, the fault induced on the Outer ring yielded the lowest results.

ACKNOWLEDGMENTS

The authors want to acknowledge the support of the Departamento Administrativo de Ciencias, Tecnología e Innovación de Colombia (Colciencias) for funding

the research project called "Development of a pilot diagnosis system for faults on rotating machines affordable to the small and medium businesses (SMB)". Code: 111052128503. We also want to thank the Mechanical Vibrations Laboratory at Universidad Tecnológica de Pereira for providing their database for this work.

REFERENCES

- [1] A. Heng, S. Zhang, A.C.C. Tan and J. Mathew. "Rotating machinery prognostics: State of the art, challenges and opportunities". *Mechanical Systems and Signal Processing*. Vol. 23, pp. 724-739. Elsevier. 2009.
- [2] S.J.Z. Sikorska, M. Hodkiewicz and L. Ma. "Prognostic modelling options for remaining useful life estimation by industry". *Mechanical Systems and Signal Processing*. Vol. 25, pp. 1803-1836. Elsevier. 2011.
- [3] K. M. Blache. "Benchmarks Shed Light on Maintenance & Reliability Perceptions, Reliabilityweb.com". 2015. Fecha de consulta: Diciembre 18 de 2015. URL: http://reliabilityweb.com/index.php/articles/benchmarking_a_better_understanding
- [4] The International Energy Agency. "KeyWorld Energy Statistics". 2013, International Energy Agency. Cedex, Paris, 2013.
- [5] J. Yan, C. Guo and X. Wang. "A dynamic multi-scale Markov model based methodology for remaining life prediction". *Mechanical Systems and Signal Processing*. Vol. 2, pp. 1364-1376. Elsevier. 2011.
- [6] F. Camci and R.B. Chinnam. "Dynamic Bayesian Networks for Machine Diagnostics: Hierarchical Hidden Markov Models vs. Competitive Learning". *Proceedings of IEEE International Joint Conference on Neural Networks*, Montreal, Canada, July 31-August 4, pp. 1752-1757. 2005.
- [7] R. Rammohan, M.R. Taha. "Exploratory Investigations for Intelligent Damage Prognosis using Hidden Markov Models". *Systems, Man and Cybernetics. 2005 IEEE International Conference on*, 12-12 Oct., Vol. 2, pp. 1524-1529, 2005.
- [8] T. Hongzhi, Z. Jianmin, J. Xisheng, J. Yunxian, Z. Xinghui and C. Liying. "Experimental Study on Gearbox Prognosis Using Total Life Vibration Analysis". 2011 Prognostics

- [9] & System Health Management Conference (PHM2011 Shenzhen), IEEE, pp. 1-6, 2011.
- [10] Z. Zhou, C. Hu, D. Xu, M. Chen, D. Zhou. "A model for real-time failure prognosis based on hidden Markov model and belief rule base". *European Journal of Operational Research* 207, Elsevier, pp. 269-283, 2010.
- [11] H. Kim, A.C.C. Tan, J. Mathew, B. Choi. "Bearing fault prognosis based on health state probability estimation". *Expert Systems with Applications*, pp. 5200-5213, 2012.
- [12] A. Widodo, B. Yang. "Machine health prognostics using survival probability and support vector machine". *Expert Systems with Applications*, Vol. 38, Elsevier, pp. 8430-8437, 2011.
- [13] Z. Zhou, C. Hu, W. Wang, B. Zhang, D. Xu, J. Zheng. "Condition-based maintenance of dynamic systems using online failure prognosis and belief rule base". *Expert Systems with Applications* 39, Elsevier, pp. 6140-6149, 2012.
- [14] E. Zio, F. Di Maio. "A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system". *Reliability Engineering and System Safety* 95, Elsevier, pp. 49-57. 2010.
- [15] R. Huang, L. Xi, X. Li, C.R. Liu, H. Qiu, J. Lee. "Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods". *Mechanical Systems and Signal Processing*, Vol. 21, pp. 193-207, 2007.
- [16] S. Sajjad, H. Zaidi, S. Aviyente, M. Salman, K. Shin, E.G. Strangas. "Prognosis of Gear Failures in DC Starter Motors Using Hidden Markov Models". *IEEE Transactions on Industrial Electronics*, Vol. 58 N° 5, pp. 1695-1705, May 2011.
- [17] J. Luo, M. Namburu, K. Pattipati, L. Qiao, M. Kawamoto and S. Chigusa. "Model-based Prognostic Techniques". Research supported by Toyota Technical Center. 2004.
- [18] J. Luo, F. Tu, M. Azam, K. Pattipati, L. Qiao and M. Kawamoto. "Intelligent model-based diagnostics for vehicle health management". *Proceedings of SPIE Conference, System Diagnosis and Prognosis: Security and Condition Monitoring Issues III*. Vol. 5107, pp. 13-26. Orlando, Florida. 2003.
- [19] C. Bunks, D. McCarthy, T. Al-Ani. "Condition-Based Maintenance of Machines Using Hidden Markov Models". *Mechanical Systems and Signal Processing*, Vol. 14 N° 4, pp. 597-612. 2000.
- [20] L. Liporace. "Maximum likelihood estimation for multivariate observations of markov sources". *Proceedings of IEEE Transactions on Information Theory IT-28*, pp. 729-734, 1982.
- [21] A. Poritz. "Linear predictive hidden markov models and the speech signal". *Proceedings of ICASSP*, pp. 1291-1294. Paris, France. 1982.
- [22] L. Rabiner. "A tutorial on hidden markov models and selected applications in speech recognition". *Proceedings of IEEE*, Vol. 77 N° 2, pp. 257-285, 1989.
- [23] H. Ocak, K.A. Loparo, "Estimation of the running speed and bearing defect frequencies of an induction motor from vibration data". *Mechanical Systems and Signal Processing*, Vol. 18, pp. 515-533, 2004.
- [24] F.V. Nelwamondo and T. Marwala. "Faults detection using Gaussian mixture models, Mel-Frequency Cepstral Coefficients and Kurtosis". *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 290-295, Taipei, Taiwan, October 8-11, 2006.
- [25] K.A. Loparo. Seeded Fault Test Data, Bearing Data Center, Case Western Reserve University. URL: <http://csegroups.case.edu/bearingdatacenter/pages/12k-drive-end-bearing-fault-data>
- [26] L. Muda, M. Begam, I. Elamvazuthi. "Voice Recognition Algorithms using Mel Frequency, Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques". *Journal of computing*. Vol. 2, Issue 3. ISSN: 2151-9617. March, 2010.
- [27] J.B. MacQueen. "Some Methods for classification and Analysis of Multivariate Observations". *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press: 1, pp. 281-297. 1967.
- [28] T. Kanungo, D.M. Mount, N.S. Netanyahu, C.D. Piatko, R. Silverman and A. Y. Wu. "An Efficient k-Means Clustering Algorithm: Analysis and Implementation". *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 24 N° 7. July 2002.

- [28] E. Yariv. "Hidden Markov Processes". *Proceedings IEEE Transactions on Information Theory*. Vol. 48 N° 6, pp. 1518-1569. June 2002.
- [29] J.C. Castaño y C.H. Agudelo. "Extensibilidad de criterios de decisión para un clasificador de fallos en rodamientos basado en HMM, sobre resultados en un clasificador de menos estados, empleando características en diferentes espacios de representación". Tesis para optar al grado de Ingeniero Electricista. Universidad Tecnológica de Pereira. T621.31042 C346; 6310000108472 F3256, 2015. URL: <http://hdl.handle.net/11059/5198>
- [30] H.A. Blandón y Y.F. Martínez. "Desarrollo de una metodología para la detección de fallas a partir del análisis de vibraciones en rodamientos usando Características de tiempo y modelos ocultos de Markov". Tesis para optar al grado de Ingeniero Electricista. Universidad Tecnológica de Pereira. T621.319 B642; 6310000106401 F2390, 2013. URL: <http://hdl.handle.net/11059/3847>
- [31] M. Shashanka. "A fast algorithm for discrete HMM training using observed transitions". United Technologies Research Center, East Hartford, CT 06108. Boston University. Available on: <http://cns.bu.edu/~mvss/stuff/ShashankaICASSP2011.pdf>
- [32] J.R. Movellan. "Tutorial on Hidden Markov Models". Machine Perception Laboratory, UC San Diego. Available on: <http://mplab.ucsd.edu/tutorials/hmm.pdf>.
- [33] J.A. Hanley and B.J. McNeil. "The meaning and use of the area under the Receiver Operating Characteristic (ROC) curve". *Radiology*. Vol. 143, pp. 29-36. 1982.
- [34] N.B. Mohamad. "Optimization of decision threshold for speaker recognition". Faculty of Electrical Engineering, University Teknologi Malaysia. October. 2003.
- [35] A.R. van Erkel and P. Pattynama. "Receiver operating characteristic (ROC) analysis: Basic principles and applications in radiology". *European Journal of Radiology*. Vol. 27, pp. 88-94. 1998.
- [36] A.J. Dwyer. "In pursuit of a piece of the ROC". *Radiology*. Vol. 202, pp. 621-625. 1997.
- [37] S. Vida. "A computer program for non-parametric receiver operating characteristic analysis". *Comput Methods Programs Biomed*. Vol. 40, pp. 95-101. 1993.
- [38] E.R. DeLong, D.M. DeLong and C.P. DL. "Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach". *Biometrics*. Vol. 44, pp. 837-845. 1988.