

# Bearings Prognostic using Mixture of Gaussians Hidden Markov Model and Support Vector Machine

F. Sloukia<sup>1</sup>, M. El Aroussi<sup>1</sup>, H. Medromi<sup>2</sup>, M. Wahbi<sup>1</sup>

<sup>1</sup>Electrical Engineering Department  
SIR2C2S/LASI-EHTP

<sup>2</sup>Laboratoire d'Informatique, Systèmes et Énergies Renouvelables  
Team Architecture of Systems, ENSEM, Hassan II University  
Casablanca, Morocco

sloukia.fatimaezzahra@gmail.com, moha387@yahoo.fr, hmedromi@yahoo.fr, wahbi\_mho@yahoo.fr

**Abstract**—Prognostic of future health state relies on the estimation of the Remaining Useful Life (RUL) of physical systems or components based on their current health state. RUL estimation can be done by using three main approaches: model-based, experience-based and data-driven approaches. This paper deals with a data-driven prognostics method which is based on the transformation of the data provided by the sensors into models that are able to characterize the behavior of the degradation of bearings.

For this purpose, we used Support Vector Machine (SVM) as modeling tool. The experiments on the recently published data base taken from the platform PRONOSTIA clearly show the superiority of the proposed approach compared to well established method in literature like Mixture of Gaussian Hidden Markov Models (MoG-HMMs).

**Keywords**—Prognostic; RUL; MoG-HMM; SVM.

## I. INTRODUCTION

The main purpose of prognostic is estimating the remaining useful life (RUL) of a failing component or subsystem so that maintenance can be executed to avoid catastrophic failures.

The choice of bearings can be explained by the fact that these components are considered as the most common mechanical elements in industry and are present in almost all industrial processes, especially in those using rotating elements and machines. Moreover bearing failure is one of the foremost causes of breakdowns in rotating machinery and such failure can be catastrophic [1], resulting in costly downtime. Many previously studies [2-4] have developed theoretical foundation and tools to describe bearing failure modes.

In this paper, data driven approach is addressed in order to utilize the availability of condition monitoring data [5]. They can be divided into statistical methods (multivariate statistical methods, linear and quadratic regression ...) and Artificial Intelligent (AI) methods which have been largely applied to machinery remaining life prediction [6]. The most used models for prognostics are artificial neural networks (ANNs) [7], support vector machine (SVM) [8], Fuzzy theory [9]. SVMs shows outstanding performance in the classification process compared with the other classifiers [10]. Also Markov chain model based on the transition probability matrix is appropriate to the analysis of a random dynamic system [11].

The proposed methods rely on two main phases: a learning phase and an exploitation phase [12]. During the first phase, the raw data are used to extract reliable features, which are then used to learn behavioral models representing the dynamic of the degradation in the bearing. In the second phase, the learned models are exploited on line to assess the current health state of the bearing and to estimate the value of the RUL.

The modeling of the degradation is done by using mixture of Gaussians Hidden Markov Model (MoG-HMM) and Support Vector Machine (SVM).

For accurate assessment of the residual life of bearings, Kim et al. [3] proposed a machine prognostics model based on health state estimation using Support Vector Machines (SVM). Kankar et al. [4] have also shown the effectiveness of SVM for bearing faults classification.

Markov chain model based on the transition probability matrix is appropriate to the analysis of a random dynamic system [11]. Li et al. Chinam and Baruah [13] have used hidden Markov models (HMM) to assess the degradations of bearings and to estimate the RUL. In their method the authors considered the degradation as a stochastic process with several states representing different health states of the physical component.

This work presents a comparison study of the performance of two well known methods: SVM and MoG-HMM.

The proposed failure prognostic methods are tested on a condition monitoring data base [5] taken from the platform PRONOSTIA [14], related to bearings degradation tests, enabling the verification of condition monitoring, fault detection, fault diagnostic and prognostic approaches.

The remainder of this paper is organized as follows. A technical background is given in Section 2 in which the MoG-HMM and SVM models and the WPD feature are discussed. Section 3 presents experiments an results in which the proposed MoG-HMM and SVM methods are validated using experimental vibration monitoring data collected from bearings. Conclusions are given in Section 4.

## II. TECHNICAL BACKGROUND

### A. Wavelet Packet Decomposition (WPD)

In the **WPD** analysis of a signal, the signal is filtered with both low-pass (LP) and high-pass (HP) filters. The LP and the HP filtered signals are referred to the approximation (A) and the detail (D), respectively. (A) and (D) are both half size of the original signal and represent the low frequency and the high-frequency content of the signal. A detailed review of WPD can be found in [15].

A third level WPD of a signal is illustrated in Fig. 1. In this representation, the third level signals AAA, DAA, ADA, DDA, AAD, DAD, ADD and DDD represent the frequency content of the original signal within the bands:  $0 - \frac{fs}{16}, \frac{fs}{16} - \frac{2fs}{16}, \frac{2fs}{16} - \frac{3fs}{16}, \frac{3fs}{16} - \frac{4fs}{16}, \frac{4fs}{16} - \frac{5fs}{16}, \frac{5fs}{16} - \frac{6fs}{16}, \frac{6fs}{16} - \frac{7fs}{16}$  and  $\frac{7fs}{16} - \frac{fs}{2}$ , respectively, where  $fs$  is the sampling rate of the signal. The energy of the signal in any layer is referred to as the node energy that we will later utilize as feature of a vibration signal.

### B. Mixture of Gaussians Hidden Markov Model (MoG-HMM)

The MoG-HMMs have proved to be a suitable tool as they model the physical component's degradation by using continuous observations provided by the monitoring sensors. They also permit the estimation of the stay durations in each health state leading to the prediction of the RUL value [16].

The hidden Markov model is defined by the parameters cited in [17]. A compact notation  $H = (A, B, \pi)$  is used for an HMM model. With  $A$  = transition probability matrix,  $B$  = observation probability matrix,  $\pi$  = initial state distribution. In practice, HMMs are used to solve typical problems [17] (The evaluation problem, the decoding problem and the learning problem).

The problem with HMM approach is that it uses discrete observations, whereas in our case we use continuous features, so changes were made to the observation probability matrix  $B$  [17], in this case we talk about MoG-HMM model. During the learning phase we used a three states left to right MoG-HMM model.

### C. Support Vector Machine (SVM)

The basic idea of SVM is to find a hyperplane which separates the  $N$  dimensional data perfectly into its two classes. However, since data is often not linearly separable, SVM introduce the notion of a "kernel induced feature space" which casts the data into a higher dimensional space where the data is separable.

Among the kernel functions in common use are linear functions, polynomials functions, gaussian basis functions (GBF) and sigmoid functions. SVMs were originally designed for binary classification and there are methods which are applicable to multiclass classification, such as "one-against-one", "one-against-all". The most suitable method is chosen according to the application constraints, the number of classes and the number of training samples [18].

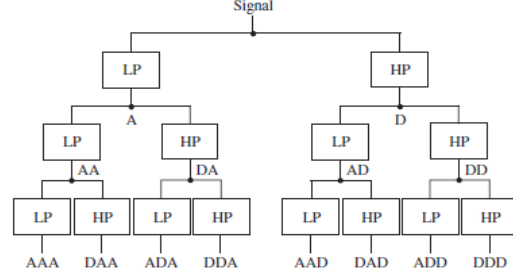


Fig. 1: Third level wavelet packet decomposition of a signal.

## III. EXPERIMENTS & USING BEARING FAILURE DATA

### A. The (WPD or RMS) and MoG-HMM prognostic method

For both learning and exploitation phases features (RMS, Crest Factor, Kurtosis, FFT, Cepstrum, WPD ...) have been extracted from raw data (vibration signal), then we have chosen two features: the RMS which is a temporal feature that detect abnormally high energy dissipation in the vibration signal [21] and the nodal energy of the WPD which has shown great results in many previous studies [16,19], it offers a richer range of possibilities for signal analysis and allows the best matched analysis to a signal. It provides level by level transformation of a signal from the time domain into the frequency domain. The choice of the 3<sup>rd</sup> level is motivated by Eq.1 where  $L$  is the decomposition level,  $F_s$  is the sampling frequency, and  $F_d$  is the defect frequency, which gives a frequency band of 3.2 KHz in each band.

$$L \leq \log_2 \frac{F_s}{3F_d} - 1 \quad (1)$$

In the **learning phase** the features extracted have been used to estimate the  $(A, B, \pi)$  parameters using the Baum-Welch algorithm and temporal parameters (the stay duration in each state) using the Viterbi algorithm of the MoG-HMM model.

Three states are defined for each MoG-HMM. The number of mixtures is fixed at two. The parameters  $(A, B, \pi)$  are firstly arbitrary initialized then the feature is used to re-estimate the parameters which are adjusted by maximizing the probability  $P(O|H)$  of the model over the observation sequence  $O$ .

Six MoG-HMMs are obtained using Baum-Welch algorithm. The RMS feature has shown better results compared to WPD because this latter need a lot of models in the same condition to give precise results. An example of estimation of the parameters of bearing 1 condition 1 is illustrated hereafter:

$$\pi = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} A = \begin{bmatrix} 0.9985 & 0.0015 & 0 \\ 0.0015 & 0.9956 & 0.0029 \\ 0 & 0.0014 & 0.9986 \end{bmatrix} M = \begin{bmatrix} 0.4176 & 0.5824 \\ 1 & 0 \\ 0.1271 & 0.8729 \end{bmatrix}$$

The Viterbi algorithm allows us to estimate the stay duration in each state and to identify the final state which corresponds in our case to the degradation state. An example of estimation of the decoded state sequence is illustrated in Fig. 2, the x-axis represents time and y-axis represents the state of the learning bearing.

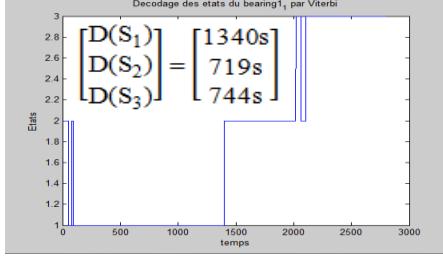


Fig.2: The decoded state sequence using Viterbi and the stay duration in each state.

The duration of healthy state of this bearing is 1340s, the average state is 719s and the faulty state is 744s. Note that the time between two measurements is 10s.

**The exploitation phase** allows to characterize the health state of test data through the selection of the model that maximizes  $P(O | H)$ , the model must be able to estimate the remaining life of the test bearing. For this we followed these steps:

- Define the failed state that corresponds to the state S3 (Faulty State),
- Estimate stay duration in each state from the path estimated by Viterbi algorithm,
- Make the sum of the lengths of stay durations,
- At every moment, we estimated the current state, at each instant, of the test bearing using the Viterbi algorithm on the test data,
- Estimate the time remaining before reaching the final state based on stay durations previously estimated.

An example of calculation of the RUL and the corresponding error is shown in Fig. 3. The x-axis represent current time of the test bearing and y-axis represent the failure time. The red line represents real RUL and the blue one represents the estimated RUL of the test bearing. We obtain: an estimated RUL of 2h 26min 10s and the relative error between the two RULs is 38.52%, with a real RUL of 3h 58min.

#### B. The (WPD or RMS) and SVM prognostic method

In **the learning phase** we form learning models based on the features (RMS and Nodal Energy of WPD) and using the LIBSVM library [20], it contains many classification modules and supports multiclass classification and cross-validation (developed later), it also allows the use of different kernels: linear, polynomial ... The resolution and performance of the SVM method involves the selection of several parameters [20]: the type of kernel, kernel parameters ( $\gamma$ , ...) and the soft-margin constant,  $C$ , which controls the penalty associated to the examples and set the relative importance of maximizing the margin and minimizing the amount of slack.

In our case we use the one against all multiclass SVM classification (3 classes: healthy average and faulty states). Because classifiers are sensitive to the way features are scaled, features must be normalized before the learning phase. The kernels chosen are polynomial and RBF kernel because they gave the most acceptable results. The cross validation method [20] is used to select ( $C$ ,  $\gamma$ ).

After constructing learning models, we pass to **the exploitation phase** in which we classify test data based on the learned models.

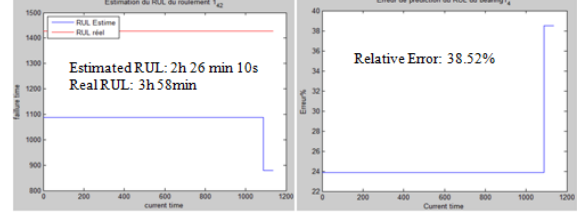


Fig 3. The estimated RUL and the corresponding relative error of the bearing 1\_4 using MoG-HMM.

For that we used `svmpredict()` function in the LIBSVM library, this function gives the percentage of classification of test data regarding the learning model (Accuracy percentage), and then we made the choice of the most suitable model.

The obtained results for the bearings 1\_3 to 1\_7 condition 1 with the model based on bearing 1\_2 applied on RMS and nodal energy of WPD features are given in TABLE I. According to the results, we find that SVM gives more accurate results using RMS than the nodal energy of WPD this is due to the fact that data provided from the WPD are correlated which degrade the performance of the SVM.

After selecting the models we use them to determine the remaining useful life of each test bearing using the state sequence of test data and by following the steps listed in MoG-HMM.

An example of calculation of the RUL and the corresponding error is shown in Fig. 4. For the same test bearing 1\_4 (real RUL: 3h 58min) we find an estimated RUL of 3h 06min and a relative error between the two RULs of 21.85%.

We note from the obtained results by using both methods (SVM and MoG-HMM) ( Figs. 3 and 4) that the estimated RUL using SVM is more closer to the real RUL than the one estimated using MoG-HMM and the relative error of this latter is higher than the one obtained using SVM. We conclude from the obtained results that the SVM presents better results compared to MoG-HMM because this latter requires more learning models to be able to give precise results and SVM is more easier to use.

TABLE I: Classification results of test data (RMS, WPD) towards the learning model based on bearing 1\_2 condition

		Accuracy en (%)				
Model based on Bearing 1_2	Bearing Feature+Parameters	1_3	1_4	1_5	1_6	1_7
	Horizontal RMS Best C=8, Best $\gamma$ =2 kernel: polynomial	99.17%	99.30%	100%	99.91%	100%
	Horizontal nodal energy of WPD Best C=2, Best $\gamma$ =0.0625 kernel: polynomial	79.58%	95%	99.09%	97.48%	99.87%

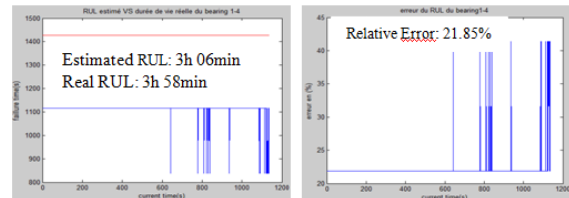


Fig.4: The estimated RUL and the corresponding relative error of the bearing 1\_4 using SVM method.

#### IV. CONCLUSION

This work was dedicated to the characterization of the health state of bearings degraded in an accelerated way by the platform PRONOSTIA, it presents a comparison of the performance of two well known methods: SVM and MoG-HMM, the choice of the MoG-HMM is motivated by its good capacity of modeling the temporal dependences which exist between the observations. It need to make an hypothesis about the probability distribution of observations in the model's states: we presuppose a particular form of data distribution. The SVM is a good classifier with a good power of generalization, but the SVM does not allow representing the temporal evolution of the observations, while this temporal evolution is essential to discriminate the vibratory signals. Our perspective thus is to exploit the advantages of SVM and HMM to construct a single hybrid method: **SVM-HMM**.

#### REFERENCES

- [1] U. Pulkkinen, S. Uryas'ev, Optimal operational strategies for an inspected component. In: Petersen KE, Rasmussen B, Eds. Safety and Reliability '92, Proceedings of the European Safety and Reliability Conference '92. London: Elsevier, pp. 896–907, 1992.
- [2] S. Marble, B.P. Morton, Predicting the remaining life of propulsion system bearings, in: Proceedings of the 2006 IEEE Aerospace Conference, Big Sky, MT, USA, 2006.
- [3] H.E. Kim, Andy C.C. Tan, J. Mathew, Eric Y. H. Kim and B.K. Choi, Machine prognostics based on health state estimation using SVM, Proceedings Third World Congress on Engineering Asset Management and Intelligent Maintenance Systems, Conference, vol. 199, pp. 834–845, 2008.
- [4] P.K. Kankar, Satish C. Sharma and S.P. Harsha, Fault Diagnosis of Ball Bearings Using Continuous Wavelet Transform, Applied Soft Computing, vol. 11, pp. 2300–2312, 2011.
- [5] IEEE PHM 2012 Prognostic Challenge, <http://www.femto-st.fr/ieee-PHM2012-data-challenge>.
- [6] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483–1510, 2006.
- [7] A.K.S. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, *Mechanical Systems and Signal Processing* 20 (2006) 1483–1510.
- [8] C. Li, T. Li, Y.S. Bai, Condition residual life evaluation by support vector machine, in: Proceedings of the Eighth International Conference on Electronic Measurement and Instruments, Xi' an, China. (2007) 441–445.
- [9] J. Fanegan, A. Ralescu, F. Ueda, A. Fuzzy, Model for estimating the remaining lifetime a diesel engine, in: Annual Meeting of the North American Fuzzy Information Processing Society, San Diego, California. (2007) 473–476.
- [10] Niu, G., Han, T., Yang, B. S., & Tan, A. C. C. (2007a). Multi-agent decision fusion for motor fault diagnosis. *Mechanical Systems and Signal Processing*, 21, 1285–1299.
- [11] T. Tien, A research on the prediction of machining accuracy by the deterministic grey dynamic model DGDM(1, 1, 1), *Applied Mathematics and Computation* 161 (2005) 923–945.
- [12] K. Medjaher, D.A. Tobon-Mejia, N. Zerhouni, "Remaining useful life estimation of critical Components with application to bearings", *IEEE TRANSACTIONS ON RELIABILITY*, VOL. 61, NO. 2, JUNE 2012.
- [13] R.B. Chinnam, P. Baruah, Autonomous diagnostics and prognostics through competitive learning driven HMM-based clustering, *Proceedings of the International Joint Conference on Neural Networks*, vol. 4, 2003, pp. 2466–2471.
- [14] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Morello, N. Zerhouni, C. Varnier, "PRONOSTIA: An Experimental Platform for Bearings Accelerated Degradation Tests" in *IEEE International Conference on Prognostics and Health Management, PHM'12.*, Denver, Colorado : United States (2012)".
- [15] C. Rowden, S. Hall, *Speech Processing*, McGraw-Hill, New York, 1992 (Chapter 5).
- [16] Diego A. Tobon-Mejia, Kamal Medjaher, Nouredine Zerhouni and Gerard Tripot , A Data-Driven Failure Prognostics Method based on Mixture of Gaussians Hidden Markov Models.
- [17] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," in *Proceedings of the IEEE*, vol.77 (2), 1989, pp. 257–286.
- [18] Jonathan Milgram, Mohamed Cheriet, Robert Sabourin, "One Against One" or "One Against All": Which One is Better for Handwriting Recognition with SVMs?
- [19] H. Ocak, K. A. Loparo, and F. M. Discenzo, "Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: A method for bearing prognostics," *Journal of sound and vibration*, vol. 302, pp. 951–961, 2007.
- [20] Chang, C-C. and Lin, C-J. (2001) LIBSVM: a library for support vector machines. Software available at <http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf>
- [21] N.G. Gebraeel, A. Elwany, J. Pan, Residual life predictions in the absence of prior degradation knowledge, *IEEE Trans. Reliab.* 58 (1) (2009).