Bayesian Approach for Remaining Useful Life Prediction

Ahmed Mosallam*, Kamal Medjaher and Nourredine Zerhouni

FEMTO-ST Institute, AS2M department, University of Franche-Comté/CNRS/ENSMM/UTBM,
24 rue Alain Savary, 25000 Besançon, France.
ahmed.mosallam@femto-st.fr

Prediction of the remaining useful life (RUL) of critical components is a non-trivial task for industrial applications. RUL can differ for similar components operating under the same conditions. Working with such problem, one needs to contend with many uncertainty sources such as system, model and sensory noise. To do that, proposed models should include such uncertainties and represent the belief about the system's state in a probabilistic form. In this work, a Bayesian approach is proposed for predicting the RUL of critical components. The approach is divided into two main parts, online and offline. In the offline part, the approach builds k-nearest neighbours classifier (kNN) for different datasets according to their end of life (EOL) values. On the other hand, the online part is similar to the offline apart from the use of Bayesian online state estimator. Bayesian online state estimator is used to represent the uncertainty of the approach about the health status. The approach starts by extracting trends that represent the health evolution of the critical component and uses these trends to build offline models of the critical component. Then, the approach uses these models to predict the RUL from new online data and assigning uncertainty value to it. The approach can be applied to a system with variable operating conditions, however, the prediction horizon will span between the minimum and maximum RUL values available in the training dataset.

1. Introduction

Maintenance strategies can be categorized into three main groups (Figure 1). In breakdown maintenance, interventions take place only when the system or the subsystem fails. In time-based preventive maintenance, interventions are placed according to periodic interval regardless of the assets health condition. Finally, in condition-based maintenance (CBM), interventions are placed according to the information collected through system condition monitoring (Jardine et. al., 2006).
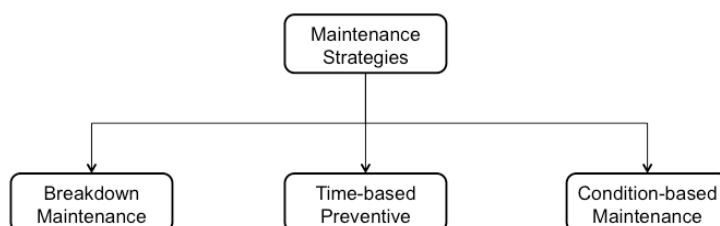

Figure 1: Maintenance strategies.

CBM strategy reduces maintenance costs while increasing efficiency by taking maintenance interventions only when there is evidence of abnormal behaviour. One of the most important CBM activities is prognostics and health management (PHM) (Medjaher et. al. 2012). In general, PHM approaches can be implemented using three different models, Physical based models, Data driven models and Knowledge based models (Schwabacher et. al.), (Figure 2).
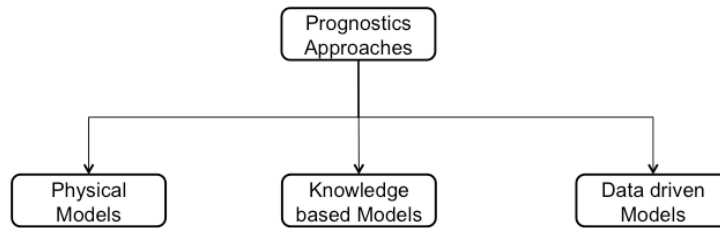
Figure 2: PHM approaches.

In the physical models, the system behaviour is characterized using state-space models, dynamic ordinary or partial differential equations and requires extensive experimentation and model verification (Luo et. al. 2003). However, once the model is built it will be very reliable until the next system upgrade.

Knowledge-based models are based on building databases of previous observed events and deducing the health status by measuring the similarity between the new observed events and the database (Biagetti et. al. 2004). Knowledge-based models can be implemented using two different approaches, expert systems and fuzzy theory.

Data-driven models are appropriate when the physical laws of the system in operation are not known. The classical data-driven models include the use of stochastic models such as the autoregressive (AR) model and the multivariate adaptive regression splines. Recently, more interests in neural networks (NNs) and neural fuzzy (NF) systems have been devoted. Different Dynamic Bayesian networks models have been used for prognostics (Medjaher et. al. 2012). These models range from semi-Markov models, Hidden and semi-hidden Markov models. Another important Bayesian estimation algorithms are Kalman and Particle filters which are not different types of models, but rather different approaches to implementing generic dynamic Bayesian networks.

PHM consists of three main routines: fault detection, diagnostics and prognostics. Prognostics has recently attracted a lot of research interest due to the need of models for accurate RUL prediction. One of the simplest forms of RUL prediction is based on trend analysis of a single monotonic parameter correlated with remaining life. This parameter may have originated from single sensor or from a number of sensors aggregated into a single variable, which is then plotted as a function of time. Although this type of trend evaluation is simple and easy to implement, there are few published examples.

In this work, an approach for data driven prognostics is presented. The approach starts by building offline trends database extracted from multidimensional datasets. These trends are later grouped according to their EOL criteria. Then, in the second stage the online data are estimated using a new discrete Bayesian filter. Finally, the RUL is predicted using kNN and the offline models.

This paper is organised as following. Section 2 outlines the method. Section 3 describes the experimental work and results. Finally a conclusion of the method work is depicted at section 4.

2. The method

The proposed approach, depicted in Figure 3, starts by looking for non-random, nonlinear and linear relationships among measured signals. The idea is that information about the wear of a component can be extracted from relations between offline signals on-board the same system (Mosallam et. al. 2011). The selected signals are then used to extract trends that represent the health status evolution through time. Next, new data acquired online are being estimated by using a non-Gaussian non-parametric Bayesian filter. The new acquired online trend is then compared to the offline trends, using kNN, to be assigned to one of the existing groups. Finally, the RUL is predicted by using the most similar offline trend curve.
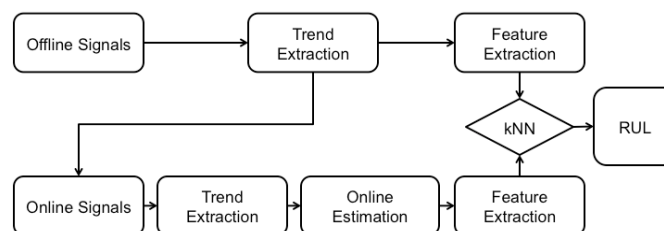

Figure 3: Overall approach scheme.

The assumptions taken in this work can be summarised as follows:
1. The input to the proposed method is multidimensional time series sensory signals.
2. The time series signals should capture the health status evolution through the time.
3. Training data sets should be available to build offline model(s).
4. The predicted RUL values will span between the values available in the offline data sets.

2.1 Offline phase

One way to build a reference model for critical components is to extract smooth monotonic trends from their sensory signals (Figure 4). The generated trends have to represent the progression of component health status and shall be used for RUL prediction. To do so we apply unsupervised trend extraction algorithm proposed by Mosallam et. al. (2012).

Figure 4: Trend extraction

Trend extraction

Not all variables measured from the critical component are useful for predicting the RUL. The assumption is that variables that have non-random relationship carry information about the system behaviour. The goal is to group the variables, which have non-random relationship. To do so, a method based on mutual information has been applied on the dataset for feature selection:

$$SU(X,Y) = 2 \times \frac{I(X,Y)}{H(X)+H(Y)}$$ (1)

Where $SU(X,Y)$ is the symmetrical uncertainty measure (Witten et. al. 2005); X and Y are two random variables; $I(X,Y)$ is the pairwise mutual information; $H(X)$ and $H(Y)$ are the information entropy for random variables (Figure 5a).

Next, in order to follow the trajectories of selected features over time, the number of features has to be reduced to a compact form. The goal in feature compression step is to compress the n features, selected in the feature selection step, onto one-dimensional space. One-way to compress the variables, i.e. reducing the number of their dimensions, without much loss of information is by using PCA. It projects data from features to principal component domain while keeping the highest variance by any projection of the n principal component. In this work standard form of principle component analysis (PCA) algorithm has been applied (Jolliffe et. al. 2002).

$$C\lambda_i = \lambda_i v_i$$ (2)

Where λ_i and v_i are eigenvalues and eigenvectors respectively and C is the covariance matrix for the selected variables is defined as:

$$C = \frac{\sum_{i}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}$$ (3)

Where $\bar{X}$ and $\bar{Y}$ denoting the mean values for X and Y, respectively. Only the first component, v_1, from the resulting PCA compression is selected to represent the health evolution (Figure 5b). However, the resulting trends are still not monotonic to be used for later modelling. To get monotonic trends, Empirical Mode Decomposition algorithm (EMD), proposed by Huang et. al (1998), has been applied to the compressed signal (Figure 5c).
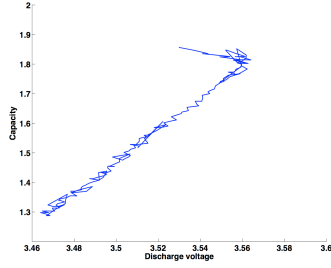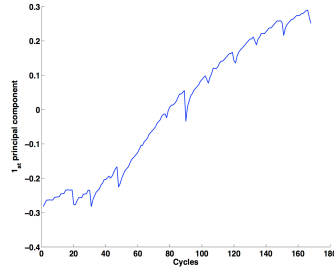
Figure 5a: Selected features.
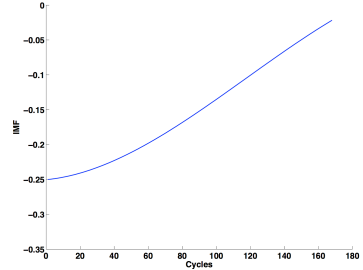

Figure 5b: First principle component.

Figure 5c: Final monotonic trend.

Figure 5: Results of the trend extraction phase.

Feature extraction

So far, trends that represent evolution of the system's health status have been extracted. These trends are used to build a reference model, and shall be used to classify new online data. In order to make the classification task more efficient, representative features should be extracted from the trends. A feature vector can represent a trend as follows:

$$< a_1(x), a_2(x), a_3(x)..... a_n(x) > \qquad (4)$$

Where x is a selected trend and a_n is the n feature extracted from the trend. In this work, two features have been extracted from each trend, linear regression coefficients and mean value. The resulting feature matrix is of size $M \times N \times Z$ where M is the length of the trend (cycles or time), N is the number of features and Z is the number of datasets used to build the offline model.

2.2 Online phase

In online phase, the same steps depicted in Figure 4 are applied to the online data. Prior to using the online trends to predict the RUL and due to noise, sensory readings might not be accurate to predict the RUL. To overcome this, sensory signals that were previously selected by the feature selection algorithm, have been estimated using online Bayesian estimator called Histogram filter, (Thrun et. al. 2005).

Histogram filter

Histogram filters decompose the state space into finitely many regions and represent the cumulative posterior for each region by a probability value. When applied to discrete spaces, such filters are called discrete Bayesian filters, and when applied to continuous spaces they are called histogram filters. In this work we assume working in a discrete space domain and therefore we apply Discrete Bayesian filter (Table 1). The input to the filter is a discrete probability distribution $\{p_{k,t}\}$, along with the recent online measurement z_t. The first line in the algorithm calculates the prediction for the new state $\overline{p}_{k,t}$,

Table 1: Discrete Bayesian filter

$$Input : \{p_{k,t-1}\}, z_t$$
$$Output : \{p_{k,t}\}$$
$$\lor x_t$$
$$\overline{p}_{k,t} = \sum_i p(X_t = x_k \mid X_{t-1} = x_i) p_{i,t-1}$$
$$p_{k,t} = \eta p(z_t \mid X_t = x_k) \overline{p}_{k,t}$$
$$end$$

based on previous state belief $p_{i,t-1}$, and state transition model $p(X_t = x_k \mid X_{t-1} = x_i)$. In the second line, the prediction is then updated by multiplying it by the measurement transition model $p(z_t \mid X_t = x_k)$, so as to incorporate the new measurement with the prediction.

Online classification

After the online estimation for the new value, the same features extracted from the offline trends are extracted from the online trends. These features are fed to kNN classifier to choose the most similar trend:

$$p(C_k \mid x) = \frac{p(x \mid C_k) \times p(C_k)}{p(x)} \qquad (5)$$

Where x is the testing pattern, C_k is the number of patterns in class k, $p(x)$ is the unconditional density associated with x, $p(C_k)$ is the class prior and $p(x \mid C_k)$ is the density associated with each class. The new data is then assigned to the class with the highest $p(C_k \mid x)$.

Distance measure
In case the online trend is assigned to a class contains more than one trend, a distance measure should be calculated to decide which trend is the most similar to the test trend. In this work, Euclidian distance is applied to measure the similarity between the testing and the offline trends.

$$\sqrt{\sum_{i=0}^{n}(q_i - p_i)^2} \tag{6}$$

Where, both q_i and p_i are test and offline trends with length of *n*.

3. Experimental work

3.1 Lithium-ion battery datasets
These data are collected on lithium-ion batteries, which ran through different operational profiles (e.g. charge, discharge and impedance) at different temperatures. Each data set, corresponding to one experiment, consists of 11 variables such as charging voltage, charging current, temperature, discharging current, discharging voltage and capacity. The ageing of the batteries was accelerated and the experiments continued until the batteries reached their EOL criteria. In this work only charge and discharge variables are used. Each cycle is presented by the mean value to reduce the processing time. Lithium-ion battery aging data set is explained in more details in (ti.arc.nasa.gov/tech/dash/pcoe/).

3.2 Results
The data sets were divided into two groups (training and testing); each group has 6 data sets (Table 2). In the offline phase the approach starts by selecting variables that have non-random relationship. One of the results of the selection algorithm is the pair {6, 11}, i.e. the voltage measured at discharge and the capacity of the battery. The approach then extracts the trends from the selected variables. The final step in the offline phase is to extract features from each trend at each cycle. The features are used as training data set and the labels are the number of cycles of each trend. For the online phase, we repeat the same steps as mentioned previously in section 2 (Figure 6). Every 10 cycles the approach classifies the input trend using a kNN classifier. Table 3 shows the results of the classification for the test data sets.

Table 2: Training and testing data sets.

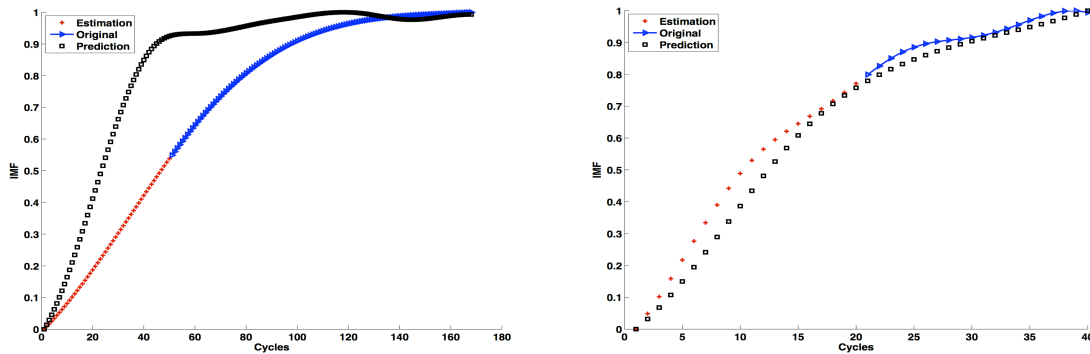| Training		Testing	
Name	Cycles	Name	Cycles
B0005	168	B0007	168
B0006	168	B0018	132
B0025	28	B0027	28
B0026	28	B0028	28
B0029	40	B0030	40
B0031	40	B0032	40

Figure 6: Prediction results for B007 (left) at cycle 50 and B0032 (right) at cycle 20.

Then the distance is measured to all trends in the same group. From the closest trend, RUL is calculated as the difference between the total number of cycles in offline trend and current cycle in the test trend. As can be seen from (Figure 7), the prediction accuracy increases when more cycles are available.

Table 3: Testing output.

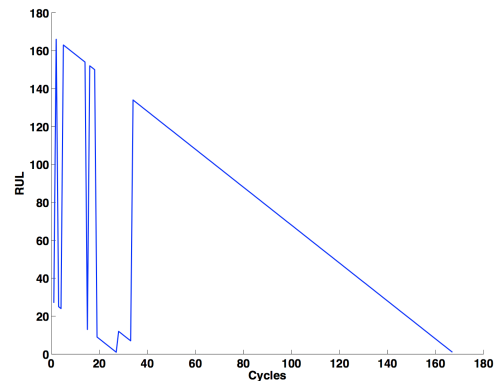| Testing		
Name	Total	Correct
B0007	17	16
B0018	14	13
B0027	3	3
B0028	3	2
B0030	4	1
B0032	4	4

Figure 7: Estimated RUL for B0007.

4. Conclusion

In this work an approach for RUL prediction has been presented. The method builds on selecting interesting variables using unsupervised feature selection algorithm. The selected features are then compressed using PCA into compact form to extract progression trends using EMD. The online data are then estimated using Histogram filters to build the online trends before extracting representative features. The similarity between the online and saved trend is being measured using Euclidian distance. The method is demonstrated on NASA datasets for batteries. The results show that the method indeed finds important relationships and used them for predicting the RUL.

References

Biagetti, T., Sciubba E., (2004). Automatic diagnostics and prognostics of energy conversion processes via knowledge-based systems, Energy, Volume 29, Issues 12–15, pp 2553-2572.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., and Liu, H.H., (1998). 'The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis'. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 454(1971), 903–995.

Jardine, A., Lin, D., and Banjevic (2006). 'A review on machinery diagnostics and prognostics implementing condition-based maintenance'. Mechanical Systems and Signal Processing 20(7):1483–1510.

Jolliffe, I. T., (2002). 'Principal Component Analysis and Factor Analysis'. In Principal Component Analysis, Springer Series in Statistics, chapter. 7, page. 150–166. Springer New York, New York.

Luo, J., Namburu , M., Pattipati , Qiao, L., K., Kawamoto, M., Chigusa, S., (2003). Model-Based Prognostic Techniques, Anaheim, CA, United States, Institute of Electrical and Electronics Engineers Inc., Piscataway, NJ, United States.

Mosallam, A., Byttner, S., Svensson, M., and T., R. (2011). Nonlinear relation mining for maintenance prediction. IEEE Aerospace Conference , 1-9. doi:10.1109/AERO.2011.5747581.

Mosallam A., Medjaher K., Zerhouni N., Eds., (2012), Unsupervised trend extraction for Prognostics and health management. 2nd IFAC workshop on Advanced Maintenance Engineering, Service and Technology, Seville, Spain.

Medjaher, K.; Tobon-Mejia, D.A.; Zerhouni, N.; , (2012). "Remaining Useful Life Estimation of Critical Components With Application to Bearings," Reliability, IEEE Transactions on , vol.61, no.2, pp.292-302.

Schwabacher, M., (2005), A survey of data-driven prognostics. AIAA InfoTech Aerospace.

Thrun ,S., Burgard, W., Fox, D., (2005). Probabilistic Robotics (Intelligent Robotics and Autonomous Agents series). Intelligent robotics and autonomous agents. The MIT Press.

Witten , H., Frank E. (2005). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, second edition.