A Fault/Anomaly System Prognosis using a Data-driven Approach considering Uncertainty

Teresa Escobet
DiPSE Department
Universitat Politècnica de Catalunya
Manresa, Spain
teresa.escobet@upc.edu

Joseba Quevedo and Vicenç Puig
ESAAI Department
Universitat Politècnica de Catalunya
Terrassa, Spain
joseba.quevedo@upc.edu; vicenc.puig@upc.edu

Abstract — This paper presents a data-driven prognostic strategy for failure prediction and computing the remaining useful life (RUL) using an autoregressive (AR) model combined with the recursive least squares (RLS) algorithm. The proposed method not only provides an estimation of the remaining useful life (RUL), but also a confidence interval based on modeling the uncertainty as a probabilistic Gaussian variable. To illustrate the performance of the proposed approach, a conveyor belt system that uses an AC electric motor to move a cart from one end to the other is used.

Keywords: prognosis, data-driven approaches, uncertainty, remaining useful life

I. INTRODUCTION

The ability to forecast machinery failure is vital to reducing maintenance costs, operation down-time and safety hazards. Occurrences of machinery failures are difficult to predict, due to the inherent structural and operational complexities of real-life systems that result from the failure interaction between components, the probabilistic nature of fault symptoms, as well as varying operating conditions and duty. Consequently, as indicated in [1], most prognostic studies have been conducted in research laboratories. In these laboratory environments, where there is frequently a lack of insight into real-life situations, it is easy to neglect certain practical considerations. Nevertheless, the ultimate goal is to establish reliable prognostic systems that can be applied in real-life situations and benefit industry.

Traditionally, prognosis has been implemented using approaches that are either model-based or data-driven [2]. The model-based approach typically involves building mathematical models to describe both the physics of the system (including the interactions between components) and physical failures, such as crack propagation or thermo-mechanical degradation. Data-driven approaches attempt to derive models directly from collected runtime data. They are based on the assumption that the statistical characteristics of the system data are relatively invariant until a fault occurs, and the anomalies, trends or patterns allow to determine a system’s state of health. In this approach, collected data are analyzed using a variety of techniques, such as statistical pattern recognition and machine-learning, which are used to detect changes in parameter data and to make predictions.

Independent of the approach used, prognostic algorithms can provide different types of outputs. Some of them estimate a health index (HI) or probability of failure at any given point, others carry out an assessment of remaining useful life (RUL), based on a predetermined failure threshold (FT). For example, in [2], the RUL is estimated using combinations of both model-based and data-driven approaches to take advantage of the strengths of each approach while overcoming their limitations. Algorithms with the ability to generate outputs that incorporate the uncertainty of predictions, such as probability distributions, fuzzy membership functions and possibility distribution, can be distinguished from others that generate only point estimates of the predictions [3]. Two active areas of research can be distinguished in the literature on prognosis and uncertainty management. The first is the development of methods that provide uncertainty bounds for output predictions and confidence levels for the values that fall within the confidence bounds. The second is uncertainty management to reduce the uncertainty bounds by using system data as more data becomes available.

This paper presents a data-driven prognostic strategy for failure prediction and computing the RUL, using an AR model combined with the RLS algorithm. The proposed method not only provides an estimation of the RUL, but also a confidence interval based on modeling the uncertainty as a probabilistic Gaussian variable. Comparing with others methodologies, this classic technique does not need knowing failure models and probability density function, because their objective is to identify trends. The paper also proposes a methodology based on a metric for tuning the parameters of the prognostic algorithm (the order of the AR model and the forgetting factor) to improve performance. To illustrate the performance of the proposed approach, a conveyor belt system that uses an AC electric motor to move a cart from one end to the other is used.

The paper is organized as follows. In Section II, uncertainty handling in prognosis methods is summarized. A method of RUL prediction is introduced in Section III. In Section IV, a conveyor belt example is used to illustrate the performance of the proposed approach. Finally, Section V summarizes the main conclusions.
II. AN OVERVIEW OF UNCERTAINTY IN PROGNOSIS

Prognosis outputs can be computed using either model-based, data-driven or hybrid approaches. In all these approaches, numerous factors affect the prognostic output and must be accounted for. These factors include measurement sensor uncertainty, model assumptions and inaccuracies, future load and usage uncertainty, loss of information due to data reduction and prediction under conditions that are different from the training data. A good prognostic system must provide an accurate and precise estimation of the RUL prediction and specify the level of confidence, considering all the uncertainties. An extended summary of the approaches used in the case of rotating machinery can be found in [1].

A. Uncertainty model-based approaches

The model-based prognostic approaches use a quantitative analytical model of the component behavior to infer the component state. These approaches also include component degradation dynamics, which can be modeled by an analytical description, given the methodology known as physics-of-failure (PoF) or a stochastic model.

However, modeling component degradations or failure evolution is not an easy task. The research that is currently conducted in this area consists in analyses to find the most appropriate methods for modeling degradation.

Physics-of-failure (PoF) is an approach that utilizes knowledge of a system’s life-cycle loading conditions, geometry, and material properties to perform reliability modeling and identify potential failure mechanisms [4]. Failure models require as input the information as material properties, geometry, environmental and operating loads. The loads are typically monitored in situ, and features (e.g., cyclic range, mean, and ramp rates) of the data are extracted and used in relevant PoF models to provide estimates of damage and RUL. The uncertainty sources are included in these models and enable to assess the impact of these uncertainties on the remaining-life distribution, in order to make risk-informed decisions. Due to the model complexity, the accumulated degradation is distributed using Monte Carlo simulations, as in [5] and [6]. From the accumulated damage distributions, the remaining life is then predicted with confidence intervals.

Failure mechanisms for which physical models have been developed include low cycle and high cycle fatigue, overstress failure, corrosion and ductile to brittle transitions.

When physical component degradation is unknown or difficult to model, statistical estimation techniques based on residuals and parity relations (the difference between the model predictions and system observations) are used to predict degradation [2]. Predictions can be improved using state estimation techniques, such as extended Kalman filters and particle filters. For example, this approach to prognostics was demonstrated for lithium ion batteries [7], where a lumped parameter model was used along with extended Kalman filter and particle filter algorithms to estimate remaining useful life (RUL).

B. Data-driven prognostics approach

The data-driven approach is considered a black box operation, since the evaluable data are used to derive models, trends or patterns, to provide information on the state of health. Data-driven approaches can be divided into two categories: statistical techniques (projection models, temporal series analysis, etc.) and artificial intelligence (AI) techniques (neural networks, fuzzy systems, etc.) [8]. These techniques use past history to infer the future and they continually update the prediction of RUL to estimate the associated uncertainty.

Statistical techniques are based on the assumption that the statistical characteristics of the system data remain relatively unchanged until a fault occurs in the system. Several approaches manage uncertainty using probabilistic representations. One of the most commonly used is the Bayesian representation, in which the predictions are represented by the corresponding probability density function. For example, in [9] a dynamic Bayesian network is used to predict the trend in the system degradation of a gas turbine compressor. In [10], a particle filter is used to predict the RUL of steam generator tubing.

With respect to AI techniques, the most promising methods are recurrent neural networks (RNNs) and neuro-fuzzy (NF) systems [8]. Only a few papers working with NN techniques include uncertainty in RUL estimation. In general, these papers combine NN with fuzzy or statistical techniques for predictions. For example, [8] propose building a library with data obtained from different transient failure scenarios (in which the residual life time is known) and estimating the RUL by matching the data evolution with the patterns in the library. RUL uncertainty is computed using a fuzzy point-wise similarity concept. In [11], an algorithm comprised of an adaptive neuro-fuzzy system and high-order particle filtering is proposed. The first forecasts the time evolution of the fault indicator and the second estimates the probability density function of RUL.

As a last remark, physic parameters do not need to be assumed in data-driven prognosis. Consequently, this technique is easy to apply. However, it needs a large amount of data to make the system as close to the real application as possible. In general, this type of method has advantages when the system is complex and no simple physical model is available.

C. Prognostic performance metrics

Independently of the approach used for prognostics, another important issue is the evaluation of prognostic performance when there are uncertainties in the estimation. As indicated in [3], performance metrics are viewed as a means of evaluating algorithm performance. Therefore, they must evaluate the prediction accuracy, taking into account the performance requirements. The performance requirements could be specified, for example, in terms of the allowable error bound (α) around the true end of life (EoL). The choice of α depends on the estimate of time required to take corrective actions. Corrective actions could be performing maintenance, bringing the system to a safe operation mode, or even, in some situations, changing control actions. Note that to do this analysis, one needs to know the true EoL of the system.
An extensive description of the metrics used to evaluate prognosis algorithms can be found in [12], in which some of them were used to evaluate uncertainty estimates. The most commonly used ([13], [3]) are: prognostic horizon and accuracy. The prognostic horizon is defined as the difference between the actual time index when the predictions first meet the specified performance criteria and the EoL. Accuracy could be evaluated using different types of metrics, but in general it is a method for comparing the actual RUL with the predicted RUL at specific time instants. It is also used to emphasize that errors closer to the actual failure of a component are more severe. Note that these metrics can only be calculated after a component has been stressed to failure. Some of them will be used in Section IV for prognostic evaluation.

III. REMAINING USEFUL LIFE PREDICTION

A. Methods for predicting the remaining life

The objectives of the prognostic methods are to extract the main features of the system; to provide one or several measures of the system’s current damage state; to track the degradation measure of the system; and to predict the remaining life. The value of degradation measure $\xi$ can be arbitrary. However, to facilitate the analysis, the scaled measure $\tilde{\xi}$ generally has values in the interval $[0,1]$, where it is assumed that failure occurs if $\tilde{\xi} \geq 1$. The RUL is then given by:

$$RUL = t(\text{end}) - t(k)$$  

(1)

where $t(\text{end})$ is the estimated time to failure for $\xi = 1$ and $t(k)$ is the current time.

A prediction of the remaining life can be obtained using three types of prior knowledge [14]: deterministic prediction, probabilistic prediction and on-line prediction.

Deterministic prediction is useful if the system will be operated according to a known sequence of mode operations (opening, closing, etc.) and mode durations. In this case, Monte Carlo simulations are performed for this “a priori” known operational sequence, and the estimated RUL (mean and variance) can be obtained from the simulations.

Probabilistic prediction could be used when the system is assumed to operate under operational sequences with a known probability. In this case, statistical methods such as particle filters are used to combine properly the probability of each sequence and to estimate the remaining life and its variance.

On-line prediction assumes that the future operation of the system will follow the observed history of the system state and the dynamic of the mode changes and will be updated from the system’s on-line information.

B. On-line prediction using autoregressive models

An autoregressive model (AR) is a parametric equation that expresses each sample as a linear combination of the previous samples plus an error term:

$$y(k) = \sum_{i=1}^{na} a_i y(k-i) + \epsilon(k)$$  

(2)

where $y(k)$ is the data $k$ of the analyzed variable, $a_i$ are model coefficients, and $\epsilon$ is white noise associated with $y$. This AR model can be written in space form [15] by

$$x(k+1) = Ax(k-1) + Ke(k)$$  

$$y(k) = Cx(k)$$  

(3)

with $na \times 1$ state vector $x(k)$ and

$$A = \begin{bmatrix} a_1 & \cdots & a_{na-1} & a_{na} \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{bmatrix}, \quad K = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}$$

The $na$ parameters of the system could be estimated using a recursive least-square (RLS) method that minimizes:

$$V_j(k) = (y(k) - \hat{y}(k))^2$$

where $\hat{y}$ is the estimated value computed by:

$$\hat{y}(k+1) = \sum_{i=0}^{na} \hat{a}_i y(k-i)$$

and $V$ is the residual variance; $\hat{a}$ denotes the estimated values of the parameters.

The estimated AR model is used for variable multi-step prediction. The $\hat{y}$ predicted on an $h$ horizon can be computed by:

$$\hat{y}(k+h|k) = C \hat{A}^h x(k),$$

(4)

The uncertainty of this multi-step prediction is computed from the $L$ given data. The residual variance of the estimation can be computed at different prediction steps $h=1,2,3…$ by

$$\hat{V}(\hat{\xi}(k+h|k)) = \sum_{p=1}^{h} C \hat{A}^p K V(k),$$

(5)

Eq. (5) allows for each $\hat{y}(k+h|k)$ to compute the standard deviation, denoted by $\sigma_k(k+h|k)$:

$$\sigma_k(k+h|k) = \sqrt{\hat{V}(\hat{\xi}(k+h|k))}.$$  

(6)

By considering a Gaussian distribution of the estimation error, a bounded envelope of the estimation of 95% can be computed on-line, using the following equations at the prediction time $h$:

$$\hat{y}(k+h|k)_{\text{max}} = \hat{y}(k+h|k) + 2\sigma_k(k)$$

$$\hat{y}(k+h|k)_{\text{min}} = \hat{y}(k+h|k) - 2\sigma_k(k).$$

(7)

The lower and upper boundary of the expected time of failure, $RUL_{\text{max}}$ and $RUL_{\text{min}}$ respectively, can be computed as follows:

$$RUL_{\text{max}} = t_{\text{max}(\text{end})} - t(k)$$

$$RUL_{\text{min}} = t_{\text{min}(\text{end})} - t(k)$$

(8)
where $t_{\text{max}}(\text{end})$ and $t_{\text{min}}(\text{end})$ are the estimated time to failure for $\xi=1$ from the upper and lower envelopes (trajectories of $\hat{y}(k+h|k_{\text{max}})$ and $\hat{y}(k+h|k_{\text{min}})$ respectively, or equivalently:

$$
t_{\text{max}}(\text{end}) = \{ [ k | \hat{y}(k+h|k_{\text{max}}) = \xi_j ] \}
$$

$$
t_{\text{min}}(\text{end}) = \{ [ k | \hat{y}(k+h|k_{\text{min}}) = \xi_j ] \},
$$

as can be seen in Fig. 1.

IV. CONVEYOR BELT HEALTH MANAGEMENT EXAMPLE

To illustrate the methodology, an example based on a conveyor belt is used.

A. Conveyor belt management problem

The conveyor belt process uses an AC electric motor to move a cart from one end to the other (Fig. 2). The motor speed is controlled by a driver that has an encoder with two input channels. The velocity set point is composed of an acceleration ramp, a constant value and a deceleration ramp. The available analog measures are: AC motor current, motor temperature, driver temperature and encoder signals. This system also has four logical sensors located in the belt and in the cart for security purposes and two control signals to move the motor to the right or to the left. The SHM has been implemented in the integrated development environment LabWindows CVI 9.0, working on a Windows platform. The system has been designed using multithreading methodology and the CompactDAQ-9172 system acquisition of National Instruments (Fig. 2).

B. Prognostic algorithm verification

The algorithm was verified in the laboratory. To evaluate the conveyor belt performance in degraded situations, several scenarios were assessed that forced the working conditions. One of them was an increase in belt friction.

Fig. 4 shows one of the features extracted from the data. It corresponds to the conveyor belt acceleration time per cycle. Notice that this feature increases abruptly at the end of process life.

The architecture proposed to estimate the $RUL$ in the case study is presented in Fig. 3. The prognostic module includes two sub-modules: an AR model estimator and an $RUL$ predictor that is part of the condition monitoring system described in [16]. The condition monitoring system allows the main features to be extracted per cycle of work. These features include motor temperature, conveyor speed at constant velocity, peaks of intensity during acceleration, acceleration and slowdown time. The prognostic module only works when new data is stored.

The architecture proposed to estimate the $RUL$ in the case study is presented in Fig. 3. The prognostic module includes two sub-modules: an AR model estimator and an $RUL$ predictor that is part of the condition monitoring system described in [16]. The condition monitoring system allows the main features to be extracted per cycle of work. These features include motor temperature, conveyor speed at constant velocity, peaks of intensity during acceleration, acceleration and slowdown time. The prognostic module only works when new data is stored.

The architecture proposed to estimate the $RUL$ in the case study is presented in Fig. 3. The prognostic module includes two sub-modules: an AR model estimator and an $RUL$ predictor that is part of the condition monitoring system described in [16]. The condition monitoring system allows the main features to be extracted per cycle of work. These features include motor temperature, conveyor speed at constant velocity, peaks of intensity during acceleration, acceleration and slowdown time. The prognostic module only works when new data is stored.
A summary of the RUL prediction compared with the actual RUL is shown in Figure 7. The red continuous line shows the actual RUL; the upper and lower RUL predictions are marked with ‘+’ and ‘o’ respectively; and the blue line indicates the predicted RUL bounds and its mean value. The accuracy and precision of the remaining life estimation using this technique improves as the end of life approaches. These results show what really happens in many mechanical devices, where in some conditions of deterioration, some features evolve very quickly towards a critical situation. In this case, the proposed methodology could predict end of life when the mechanical device is near to its end.

This method has been used to predict the temperature of the motor. In the conveyor belt case study, temperature is a critical variable since it is managed to keep its value below some pre-established threshold, in order to increase the system’s lifetime. If the temperature moves quickly towards a predetermined value, corrective action may be taken to prevent this.

Figure 8 shows the temperature evolution and its prediction starting at cycle 41. In contrast to the previous case, the temperature evolves slowly over time. The pre-established threshold is 65% of the critical value.

Figure 9 shows the estimated and the actual RUL. It can be seen that the proposed method applied to temperature evolution allows us to predict the evolution of the RUL.
Figure 9. Illustration showing the estimated RUL of the temperature.

C. Prognostic performance evaluation

As seen in Section III and its application above, the proposed algorithm has two parameters that should be set previously: the AR model order and the forgetting factor value. These two parameters have been tuned off-line by comparing prognostic performance using a metric. The metric used in this comparison is the prognostic horizon. Taking into account the definition proposed in [12], the prognostic horizon, \( PH \), is computed as follows:

\[
PH = t_{EoL} - i, \quad (10)
\]

where:

- \( i = \min \left\{ j | (j \in h) \land \left( R_{RUL,j} \subset R_k^i \right) \quad \forall k \in [j, t_{EoL}] \right\} \) is the first time index when predictions satisfy \( \alpha \)-bounds and it remains inside;
- \( h \) is the set of all time indexes when a prediction is made;
- \( R_{RUL,j} \in [RUL_{j,\text{min}}, RUL_{j,\text{max}}] \), is the predicted RUL at time \( j \) and \( [RUL_{j,\text{min}}, RUL_{j,\text{max}}] \) is the confidence interval.
- \( t_{EoL} \) is the actual end of life;
- \( R_k^i = [t_{EoL}(1 - \alpha) - i(k), t_{EoL}(1 + \alpha) - i(k)] \) are the \( \alpha \)-bounds of the actual RUL;
- and, \( \alpha \) is the allowable error bound of the actual EoL.

Some authors, as in [17], introduce the concept of the percentage of the probability mass function that falls within the \( \alpha \) bounds and gives the \( PH \) associated with a probability. In this case, the \( PH \) is considered inside \( \alpha \) bounds only if the probability mass function is above a predetermined threshold. The metric proposed in Eq. (10) is more conservative than the previous one. It aims to ensure that near EoL, the \( RUL \) prediction remains within the \( \alpha \) bounds and reduces the spread of information provided by the prognostic module.

Figure 10 displays with a yellow band the desired level of accuracy, \( R_k^i \), with respect to the EoL and specified by the allowable error bound of \( \alpha = 2.5\% \). Magenta and blue lines draw the predicted RUL bounds and the mean value arising from using the RLS with a forgetting factor of 0.92 and 0.90, respectively. With a confidence level of 95\%, the prediction horizon is \( PH_1 = 9 \) in the first case and \( PH_2 = 11 \) in the second. The \( PH \) is obtained when the uncertainty RUL prediction is included in the desired accuracy zone.

Figure 11 displays results from the same example presented in Figure 10, but in the case that the forgetting factor is 0.85. Using the metric described in Eq. (10), \( PH_1 = 10 \), and only taking into account the first case, the predictions that satisfy \( \alpha \)-bounds \( PH_3 \) will be 17. Thus, the prognostic horizon metric could help the module predictor designer to select the parameters that enable better prediction of the system RUL. As discussed above, in the application example used in this paper, the \( PH \) metric was used to select both the forgetting factor and the AR order.

Figure 11. Illustration showing the estimated RUL with \( \alpha \) error bound and RLS with a forgetting factor of 0.85 and 0.90.
V. CONCLUSIONS

This paper proposed a data-driven prognostic strategy for failure prediction and computing the remaining useful life (RUL) on-line using an AR model combined with the RLS algorithm. The proposed method estimates not only the RUL, but also a confidence interval based on modeling the uncertainty as a probabilistic Gaussian variable. The paper also proposed a methodology based on a metric (the prognostic horizon) for tuning the parameters of the prognostic algorithm (the order of the AR model and the forgetting factor) to improve performance. The performance of the proposed approach was illustrated by a conveyor belt system that uses an AC electric motor to move a cart from one end to the other.

ACKNOWLEDGMENTS

This study was funded by the Spanish Ministry of Science and Technology, through the project CYCYT SHERECS MINECO DPI2011-26243, and by the European Commission through the i-Sense FP7-ICT-2009-6-270428 contract.

REFERENCES


