Production Planning & Control: The Management of Operations
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tppc20

A prognostic algorithm for machine performance assessment and its application

Jihong Yan , Muammer Koç & Jay Lee
[a] Center for Intelligent Maintenance Systems , University of Wisconsin-Milwaukee , Milwaukee, WI 53211 USA
[b] Department of Mechanical Engineering at the University of Michigan , Ann Arbor, MI 48109, USA
[c] Center for Intelligent Maintenance Systems , University of Wisconsin-Milwaukee , Milwaukee, WI 53211 USA E-mail:
Published online: 21 Feb 2007.

Production Planning & Control,
Vol. 15, No. 8, December 2004, 796–801

A prognostic algorithm for machine performance assessment and its application

JIHONG YAN, MUAMMER KOÇ and JAY LEE

Keywords Logistic regression, performance assessment, remaining life prediction, elevator door system

Abstract. This paper explores a method to assess assets performance and predict the remaining useful life, which would lead to proactive maintenance processes to minimize downtime of machinery and production in various industries, thus increasing efficiency of operations and manufacturing. At first, a performance model is established by taking advantage of logistic regression analysis with maximum-likelihood technique. Two kinds of application situations, with or without enough historical data, are discussed in detail. Then, real-time performance is evaluated by inputting features of online data to the logistic model. Finally, the remaining life is estimated using an ARMA model based on machine performance history; degradation predictions are also upgraded dynamically. The results such as current machine running condition and the remaining useful life, are output to the maintenance decision module to determine a window of appropriate maintenance before the machine fails. An application of the method on an elevator door motion system is demonstrated.

1. Introduction

Equipment degradation and unexpected failures impact the three key elements of competitiveness – quality, cost and productivity. Well-maintained machines and products hold tolerances better, help reduce downtime and rework, and increase consistency and overall business efficiency. Since generally machines go through degradation before failure occurs, monitoring the trend of machine degradation and assessing performance allows the degraded behaviour or faults to be corrected before

Authors: Jihong Yan (corresponding author) and Jay Lee, Center for Intelligent Maintenance Systems, University of Wisconsin-Milwaukee, Milwaukee, WI 53211 USA. E-mail: yanjh@uwm.edu. Muammer Koç, Department of Mechanical Engineering at the University of Michigan, Ann Arbor, MI 48109,USA.

JIHONG YAN is a postdoctoral research associate at the Center for IMS, Department of Industrial and Manufacturing Engineering, University of Wisconsin-Milwaukee, USA. Her research interests include intelligent methods for predictive maintenance, optimization scheduling methods, concurrent engineering, systems modelling, simulation and intelligent algorithms.

MUAMMER KOÇ, is an assistant research scientist in the Department of Mechanical Engineering at the University of Michigan, Ann Arbor, USA. He has been conducting research at the National Science Foundation Engineering Research Center for Reconfigurable Manufacturing Systems, the National Science Foundation Industry/University Cooperative Research Center for Intelligent Maintenance Systems, and the S.M. Wu Manufacturing Center. His research focus has been on advanced manufacturing processes, systems, forming of lightweight materials, micro-meso scale forming, e-Manufacturing and predictive modelling.

Production Planning & Control ISSN 0953–7287 print/ISSN 1366–5871 online © 2004 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/09537280412331309208

JAY LEE is Wisconsin Distinguished Professor and Rockwell Automation Professor at the University of Wisconsin-Milwaukee, director of the Center for IMS. His current research is in the areas of intelligent maintenance and self-maintenance systems. He has pioneered Watchdog Agent™ embedded prognostics technologies and web-enabled Device-to-Business (D2B)™ platform for predictive machine degradation assessment, remote monitoring and prognostics. He is a Fellow of SME and also a Fellow of ASME.

they cause failure and machine breakdowns (Mobley 1989, Lee 1995).

Many efforts have been made to develop methods and tools to diagnose failures (Greitzer *et al.* 1999, Kacprzynski and Roemer 2000, Roemer and Kacprzynski 2000). For machine prognostics (especially predicting the degradation), however, not much progress has been shown, though it still stands as a major barrier and opportunity for achieving intelligent manufacturing and total business efficiency. The essence of prognostics is the estimation of remaining life in meaningful terms that would lead to a profound and intelligent maintenance decision process (Swanson 2001), which would lead to proactive maintenance processes to minimize downtime of machinery and production in various industries, thus increasing efficiency of operations and manufacturing. Proactive maintenance makes it unambiguous when, where and by whom problem-solving is necessary before failure really occurs (Spear 2002).

This paper presents a prognostic method for machine degradation detection, which can both assess machine performance and predict the remaining useful life. In terms of preventive maintenance records, the machine running condition is a dichotomous problem, either normal or failure. Results in the literature indicate that analysis of dichotomous data should be conducted using the logistic regression function (Hosmer and Lemeshow 1989, Spezzaferro 1996). Logistic regression is widely used to model the outcomes of a categorical dependent variable. For categorical variables, it is inappropriate to use linear regression because the response values are not measured on a ratio scale and the error terms are not normally distributed (Czepiel, n.d.). Currently, logistic regression analysis has been used extensively in medical research to classify health condition (health or disease) (Rego and de Souza 2002). In this context, we use logistic regression to map the classification of machine running conditions from normal to failure. Based on the logistic model after training, the performance of a machine can be calculated at each calculation cycle and then, according to the previous performance assessment results, future performance tendency is predicted by an ARMA model; consequently time to failure is delivered dynamically.

In section 2, the performance model is set up using a logistic function based on training data. Also, two kinds of practical circumstances such as with or without historical data are considered. In section 3, by taking advantage of the ARMA model, the remaining useful life is predicted as the uncertain duration between the present and the point where a component can no longer perform its function. The method has been applied to an elevator door motion system performance assessment: the application results are illustrated in section 4. Finally, conclusions and recommended future work are presented in section 5.

2. Machine performance assessment

In this section, a performance assessment model using logistic regression is investigated. Two kinds of situations that historical maintenance data are enough and not sufficient are considered in this part.

Logistic regression is a technique for analysing problems where there are one or more independent variables that determine an outcome that is measured with a dichotomous variable in which there are only two possible outcomes. The goal of logistic regression is to find the best fitting model to describe the relationship between the dichotomous characteristic of the dependent variable and a set of independent variables. Here, logistic regression is used to set up the relationship between normal and failure running conditions.

2.1. *Using logistic regression analysis with historical data*

In the logistic regression method, the dependent variable is the probability that an event will occur, hence output is constrained between 0 and 1 (see equation (1)). Logistic regression has the additional advantage that all of the independents can be binary, a mixture of categorical and continuous or just continuous. The logistic function is:

$$\text{Prob(event)} = P(\vec{x}) = \frac{1}{1+e^{-g(\vec{x})}} = \frac{e^{g(\vec{x})}}{1+e^{g(\vec{x})}} \qquad (1)$$

where $\vec{x}(x_1, x_2, \ldots, x_k)$ is an input vector, corresponding to the independent variables, and $g(\vec{\boldsymbol{x}})$is the logit model.

Because $P(\vec{\boldsymbol{x}})$ ranges between 0 (normal) and 1 (failure), the logistic function can be thought of as a probability distribution function (Kleinbaum 1994). The difference in logistic regression analysis is that the outcome value has a discrete number of responses, in this case binary 0 (normal) or 1 (failure), rather than continuous. The range of the conditional mean $E(Y|x_1, x_2, \ldots, x_k)$, the expected value of Y depending on the values of $x_1, x_2, \ldots, x_k$, is between 0 (normal) and 1 (failure) for dichotomous data. The logit model is:

$$g(\vec{\boldsymbol{x}}) = \log\left(\frac{P(\vec{\boldsymbol{x}})}{1 - P(\vec{\boldsymbol{x}})}\right) = \sum_{i=0}^{k} x_i \beta_i, \quad i = 1,\ 2, \ldots, k \quad (2)$$

where $x_0 = 1$. The input array of independent variables, $(1, \vec{\boldsymbol{x}})$, is composed of $k+1$ columns, where k is the number of independent variables specified in the model. The parameter vector, $\boldsymbol{\beta}$, is a column vector of length $k+1$. There is one parameter corresponding to each of the k columns of independent variables in $\vec{\boldsymbol{x}}$, plus one, β_0, for the intercept.

The goal of logistic regression is to estimate the $k+1$ unknown parameters $\boldsymbol{\beta}$ in equation (2). For logistic regression, least squares estimation is not capable of producing minimum variance unbiased estimators for the actual parameters. In this case, maximum-likelihood estimation is used to solve for the parameters that best fit the data. This is done with maximum likelihood estimation which entails finding the set of parameters for which the probability of observed data is greatest (Czepiel, n.d.).

The approach is absolutely feasible when we have enough maintenance records including both normal and failure data to train the model. But there is usually a lack of empirical data on which prognostic calculations can be based, and tests are very difficult and expensive to perform. Consequently, there is not enough historical failure data available. Therefore, solving parameters of β without enough historical data is a challenge.

2.2. *Using logistic regression analysis without enough historical data*

When the machine is running in a satisfactory condition, we know the normal running level although there is not enough historical data available; therefore, when the machine is new or running in a stable condition, the corresponding features or inputs $(\vec{\boldsymbol{x}})$ can be acquired. But only normal level is not enough for regression analysis. Here, we take advantage of the technician's experience, and sample different inputs $(\vec{\boldsymbol{x}})$ that correspond to different running levels, such as acceptable level, unacceptable level and so on. The purpose here is to solve the parameters of $\boldsymbol{\beta}$ in equation (2) and determine the logistic model.

For the inputs of each level, $\vec{\boldsymbol{x}}$, the corresponding performance level is set in advance; for example, set the probability of failure with 'very normal level' as 0.02 (i.e. $\Pr\{\vec{\boldsymbol{x}} \in failure|\vec{\boldsymbol{x}}\} = 0.02$), and the probability of failure with 'unacceptable level' as 0.5. We can utilize the human experience and observations to obtain other levels; as long as these levels are ascertained, we can also use the logistic regression method to implement regression of the categorical problem. The maximum-likelihood method can also be used to solve the parameters.

3. Estimation of remaining useful life

In this section, prediction of degraded performance is accomplished by trending results from a logistic regression classifier module. Here, we make use of the ARMA$(p,\ q)$ model.

The ARMA$[p, q]$ or Box–Jenkins model is one of the most traditional techniques in statistical time-series analysis. The assumed model is of the form:

$$x_t = \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \cdots - \beta_q \varepsilon_{t-q} \quad (3)$$

where p is the order of the autoregressive part, q is the order of the moving-average part, $\alpha_1, \ldots, \alpha_p$ are the autoregressive parameters, and $\beta_1, \ldots, \beta_q$ are the moving-average parameters; ε_t denotes the series of errors.

The way of approaching the modelling of such an ARMA $[p, q]$ process is to first determine the model orders p and q. This part is done offline. The specific procedure is as follows:

(1) Performance calculation using equation (1) (shown in section 2) a series of performance indices is obtained.
(2) Choose suitable p, q :
    Do {
        ARMA model order $(p,\ q)$ selection
        ARMA model validation
    } while ARMA model is rejected;

After p and q are determined (Yan *et al.* 1999), the online ARMA model parameters identification is implemented using a real-life performance index; prediction can then be done based on the built dynamic ARMA model.

4. Application

4.1. *With enough historical data*

For an elevator door motion system, an encoder was installed to measure door displacement and parallel connections were made to acquire digital signals (door commands) from the door control system (see figure 1). Their readings are taken periodically.

Two quantitative variables (*open_cycle time* and *max_angular speed*) are extracted as features (relative inputs) of the door motion system for simplification and illustration. Table 1 shows the features extracted from selected [*normal*, *failure*] records.

Here, x_1 is *open_cycle time*, x_2 is *max_angular speed* corresponding to equation (2). Based on the features of normal and failure sets, the maximum likelihood technique is applied to solve the parameter vector $\boldsymbol{\beta}$. Therefore, the logit model shown in equation (2) is determined:

$$g(\vec{\boldsymbol{x}}) = \log\left(\frac{P(\vec{\boldsymbol{x}})}{1 - P(\vec{\boldsymbol{x}})}\right) = -1.4768 - 16.9629 * cycle\_time + 9.5603 * max\_speed$$

At each sampling point, feature set {*cycle_time*, max *_angular_speed*} is extracted from raw data such as displacement from the encoder and four control commands. According to the above equation, $g(\vec{\boldsymbol{x}})$ can be calculated; consequently, probability of failure at each sampling point is calculated according to equation (1). Finally, based on results of probability of failure, prediction is made according to section 3. figure 2 shows the daily average open cycle lead time over 400 days. figure 3 illustrates the change of average maximum angular speed, probability of failure at each time point is calculated by equation (1) and remaining useful life estimation is shown in figure 4.

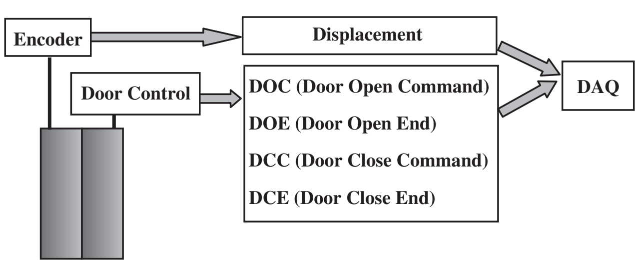

Figure 1. Door data acquisition system.

Figure 4 shows remaining useful life before maintenance is predicted dynamically. These windows also indicate the recommended maintenance windows before failure occurs. In this case, a predictive maintenance was done on the 200th day to keep the elevator door running continuously and achieve near-zero-downtime running. After maintenance, the system performance recovers.

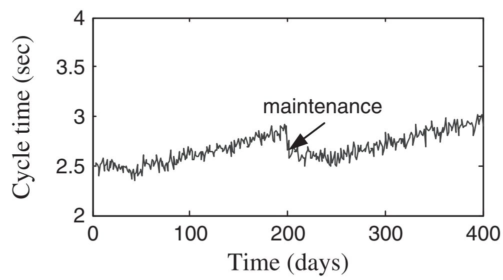

Figure 2. Cycle time.

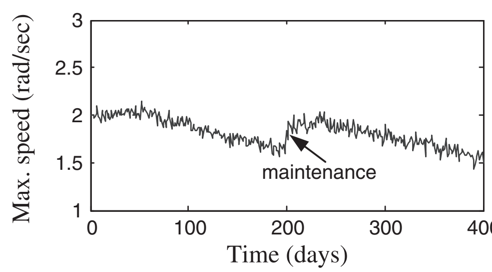

Figure 3. Maximum angular speed.

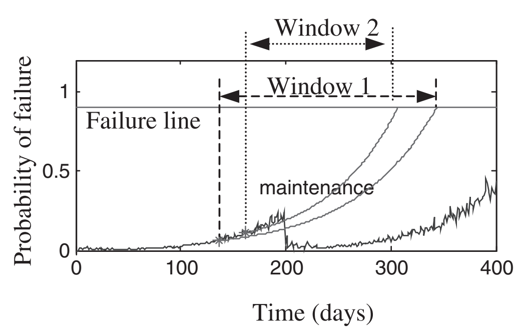

Figure 4. Failure probability calculation and remaining useful life prediction.

Table 1. Features extracted from selected [*normal*, *failure*] sets.

	Normal (0)					Failure (1)				
Cycle time (sec)	2.59	2.68	2.85	2.93	3.01	3.16	3.28	3.35	3.41	3.48
Max_speed (rad/sec)	2.06	2.03	1.96	1.91	1.77	1.31	1.24	1.17	1.04	0.99

4.2. *Without historical data*

We are also dealing with an elevator door newly installed at our centre's lab; there are not any useful historical records for this door system. The normal running condition is measured via an encoder installed on the car door by measuring door displacement. Consequently, maximum speed and open cycle time of each open-close cycle can be calculated from the sampling data. The method described in subsection 2.2 is applied to calculate probability of failure by the following levels; here, $\vec{\boldsymbol{x}} = \{open\_cycle\_time,\ max\_\ angular\_speed\}$:

(1) Normal: $\Pr\{\vec{\boldsymbol{x}} \in failure|\vec{\boldsymbol{x}}\} = 0.01$; $\vec{\boldsymbol{x}}$ is acquired from elevator door normal running conditions.

$$\left.\begin{array}{l}(2)\ \text{Acceptable}: \\ \quad \Pr\{\vec{\boldsymbol{x}} \in failure|\vec{\boldsymbol{x}}\} = 0.25; \\ (3)\ \text{Unacceptable}: \\ \quad \Pr\{\vec{\boldsymbol{x}} \in failure|\vec{\boldsymbol{x}}\} = 0.5\end{array}\right\} \begin{array}{l}\vec{\boldsymbol{x}}\ \text{is determined based} \\ \text{on the technician's} \\ \text{experiences and} \\ \text{statistical analysis} \\ \text{on sampling data.}\end{array}$$

The parameter vector $\boldsymbol{\beta}$ is also determined by the maximum-likelihood technique. Here, 376 running cycles are taken. Figure 5 shows the open cycle lead-time, the x-axis stands for the number of cycles; figure 6 illustrates the maximum angular speed of each cycle; figure 7 shows the corresponding failure probability curve.

While the door is running in normal conditions, obviously the performance value (probability of failure) is fairly stable. When some kind of disturbance appears, for example, at the 192th cycle, *max_speed* drops (figure 6), then the failure probability value increases correspondingly (figure 7). In the elevator door motion system, there are two door layers—car door and landing door. When the door opens or closes, the two layers hook up and move together. In actual life, when the elevator door is opened on different floors, even though the car door is the same, the landing door is different. Thus, when the door opens or closes on a different floor, the instant friction or misalignment may be different, this might be the reason that causes *max_speed* to drop. At each open_close cycle, the performance index is given out. When the continuous degradation occurs, future machine behaviour can be predicted by the ARMA model.

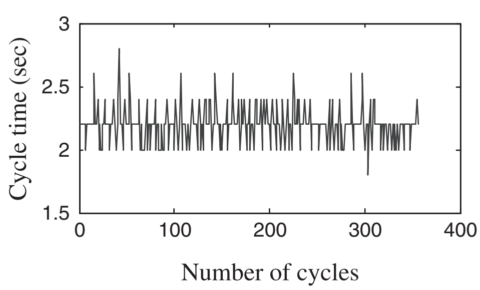

Figure 5. Open cycle lead-time.

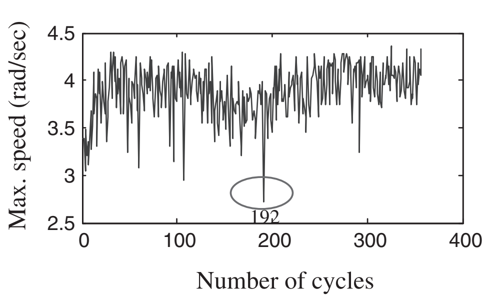

Figure 6. Maximum angular speed.

5. Conclusions

The approach presented in this paper involves three steps: (1) setting the mapping between inputs and probability of failure using logistic regression function; (2) calculating the real-time performance by the logit model; and (3) updating the equipment degradation prediction continuously, and estimating the remaining life at the same time. By this approach, the health of a component is answered at any point in time and the future failure event can be safely predicted in advance for proactive maintenance purposes.

The first version of the predictive degradation detection system introduced in this paper has been implemented for the elevator door motion system in online performance monitoring. Preliminary results for machine/system performance assessment and TTF (time to failure) prediction are very promising. Additional data such as current, vibration and delay time between command and action will be collected from the field for further development and implementation of the method. A knowledge-based fault recognition system is one of the subjects of future studies. Its output

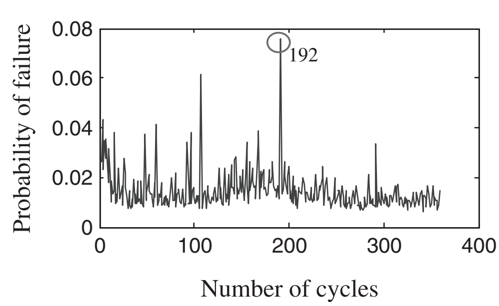

Figure 7. Probability of failure.

with TTF can be integrated with the maintenance decision to obtain optimal maintenance strategy.

References

CZEPIEL, S. A., Maximum likelihood estimation of logistic regression models: theory and implementation. Available at http://czep.net/stat/mlelr.pdf.

GREITZER, F. L., STAHLMAN, E. J., FERRYMAN, T. A. *et al.*, 1999, Development of a framework for prediction life of mechanical systems: life extension analysis and prognostics. *SOLE 1999 Symposium*, Las Vegas, NV, 30 August to 2 September.

HOSMER, D. W., and LEMESHOW, S., 1989, *Applied Logistic Regression* (Wiley).

KACPRZYNSKI, G. J., and ROEMER, M. J., 2000, Health management strategies for 21st century condition-based maintenance systems. *International COMADEM Congress*, Houston, TX, December.

KLEINBAUM, D. G., 1994, *Logistic Regression* (Springer-Verlag).

LEE, J., 1995, Machine performance monitoring and proactive maintenance in computer-integrated manufacturing: review and perspective. *International Journal of Computer Integrated Manufacturing*, **8**(5), 370–380.

MOBLEY, R. K., 1989, *An Introduction to Predictive Maintenance* (Van Nostrand Reinhold).

REGO, L. C., and MENEZES CAMPELLO DE SOUZA, F., 2002, Improved estimation of left ventricular hypertrophy. *IEEE Engineering in Medicine and Biology Magazine*, **21**(1), 66–73.

ROEMER, M. J., and KACPRZYNSKI, G. J., 2000, Advanced diagnostics and prognostics for gas turbine engine risk assessment. *Aerospace Conference Proceedings, IEEE*, **6**, 345–353.

SPEAR S. J., 2002, The essence of just-in-time: embedding diagnostic tests in work-systems to achieve operational excellence. *Production Planning & Control*, **13**(8), 754–767

SPEZZAFERRO, K. E., 1996, Applying logistic regression to maintenance data to establish inspection intervals. *Proceedings of Annual Reliability and Maintainability Symposium*, pp. 296–300.

SWANSON, D. C., 2001, A general prognostic tracking algorithm for predictive maintenance. *Aerospace Conference Proceedings, IEEE*, **6**, 2971–2977.

YAN, J. H., SHI, X. P., and WANG, Z. C., 1999, A new on-line identification method and its application to simulation of turbogenerator sealing oil cooling system. *Journal of Harbin Institute of Technology*, **31**(1), 13–16.