Data-driven approach to machine condition prognosis using least square regression trees

Van Tung Tran[†], Bo-Suk Yang[†], Myung-Suck Oh[†]

ABSTRACT

Machine fault prognosis techniques have been considered profoundly in the recent time due to their profit for reducing unexpected faults or unscheduled maintenance. With those techniques, the working conditions of components, the trending of fault propagation, and the time-to-failure are forecasted precisely before they reach the failure thresholds. In this work, we propose an approach of Least Square Regression Tree (LSRT), which is an extension of the Classification and Regression Tree (CART), in association with one-step-ahead prediction of time-series forecasting technique to predict the future conditions of machines. In this technique, the number of available observations is firstly determined by using Cao's method and LSRT is employed as prognosis system in the next step. The proposed approach is evaluated by real data of low methane compressor. Furthermore, the comparison between the predicted results of CART and LSRT are carried out to prove the accuracy. The predicted results show that LSRT offers a potential for machine condition prognosis.

Key Words: Least square method; Embedding dimension; Regression trees; Prognosis; Time-series forecasting

1. Introduction

Most of the components in machine are degraded condition during operation due to wear which is the major reason causing machine breakdown. Maintenance is the set of activities performed on a machine to sustain it on operable condition. The most common maintenance strategy is the corrective maintenance which almost means "fix it when it breaks". However, this strategy reduces the availability of machine and high unscheduled downtime. Condition-based maintenance (CBM) which involves diagnostic module and prognostic module is an alternative. Prognosis is the ability to access the current state, forecast the future state, and predict accurately the time-to-failure or the remaining useful life (RUL) of a failing components or subsystems. RUL is the time left for the normal operation of machine before the breakdown occurs or machine condition reaches the critical failure threshold.

Prognosis is a relatively new area and becomes a significant part of CBM [1]. Various approaches to prognosis have been developed that range in fidelity from simple historical failure rate models to high-fidelity physics-based models. Fig. 1 illustrates the hierarchy of potential prognostic approaches related to their applicability and relative accuracy as well as their complexity. Each of those approaches has advantages and limitations in application. For example, experience-based prognosis is the least complex, however, it is only utilized in situations where the prognostic model is not warranted due to low failure occurrence rate; trend-based prognosis may be implemented on the subsystems with slow degradation type faults [2].

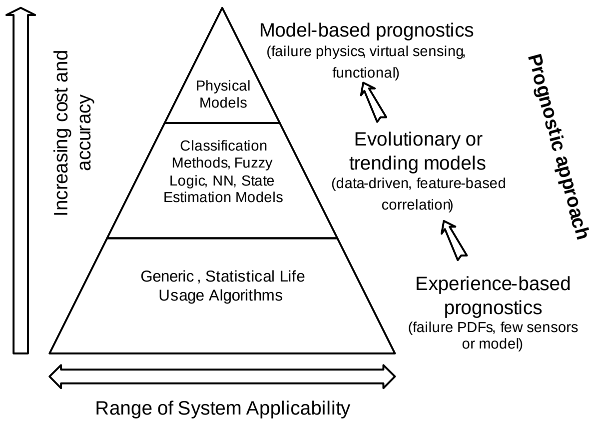

Fig. 1 Fidelity of prognostic approaches

Data-driven and model-based techniques are much considered due to their accuracy. Nevertheless, model-based techniques require accurate mathematical models of failure modes and are merely applied in some specific components in which each of them needs different model. Furthermore, a suitable model is also difficult to establish to mimic the real life. Meanwhile, data-driven techniques can generate the flexible and appropriate models for almost failure modes. Consequently, data-driven approaches are firstly examined that some of those have been proposed [3-6].

In order to predict the condition of machines, one-step-ahead or multi-step-ahead predictions of time-series

† School of Mechanical Eng., Pukyong National University
E-mail : bsyang@pknu.ac.kr
Tel : (051) 620-1604, Fax : (051) 620-1405

forecasting techniques is frequently used. They imply that the prognostic system utilizes available observations to forecast one value or multiple values at the definite future time. The more the steps ahead are, the less reliable is the forecasting operation because multi-step prediction is associated with multiple one-step operations [6].

In data-driven approaches, the number of essential observations, so-called embedding dimension d, is used for forecasting the future value. It should be chosen large enough so that the estimator can forecast accurately the future value and not too large to avoid the unnecessary increase in computational complexity. False nearest neighbor method (FNN) [7] and Cao's method [8] are commonly used to determine the embedding dimension. However, FNN method not only depends on chosen parameters and the number of available observations but also is sensitive to additional noise. Cao's method overcomes the shortcomings of the FNN approach and therefore, it is chosen in this study.

The CART [9] is widely implemented in machine fault diagnosis. In the prediction techniques, CART is also applied to forecast the short-term load of the power system [10]. Nevertheless, the average value of samples in each terminal node used as predicted result is the reason for reducing the accuracy of CART. Several approaches have been proposed to ameliorate that CART's limitation [12-14]. In this article, we suggest the use of LSRT which is an extension of the CART as an estimator for predicting the conditions of machine.

2. Background knowledge

2.1 Determine the embedding dimension

Assuming a time-series of $x_1, x_2, \ldots, x_N$. The time delay vector is defined as follows:

$$y_{i(d)} = [x_i, x_{i+\tau}, x_{i+2\tau} \ldots, x_{i+(d-1)\tau}] \quad i = 1,2,\ldots, N-(d-1)\tau \tag{1}$$

where τ is the time delay. Defining the quantity as follows:

$$a(i,d) = \frac{\left\| y_i(d+1) - y_{n(i,d)}(d+1) \right\|}{\left\| y_i(d) - y_{n(i,d)}(d) \right\|} \tag{2}$$

where $\|\cdot\|$ is the Euclidian distance and is given by the maximum norm, $y_i(d)$ means the ith reconstructed vector and $n(i, d)$ is an integer such that $y_{n(i,d)}(d)$ is the nearest neighbor of $y_i(d)$ in the embedding dimension d. In order to avoid the problems encountered in FNN method, the new quantity is defined as the mean value of all a(i, d)'s:

$$E(d) = \frac{1}{N-d\tau} \sum_{i=1}^{N-d\tau} a(i,d) \tag{3}$$

$E(d)$ is dependent on only the dimension d and the time delay τ. To investigate its variation from d to d+1, the parameter E_1 is given by

$$E_1(d) = \frac{E(d+1)}{E(d)} \tag{4}$$

By increasing the value of d, the value $E_1(d)$ is also increased and it stops when the time series comes from a deterministic process. If a plateau is observed for $d \geq d_0$, $d_0 + 1$ is the minimum embedding dimension.

The Cao's method also introduced another quantity $E_2(d)$ in case that $E_1(d)$ is slowly increasing or has stopped changing if d is large enough:

$$E_2(d) = \frac{E^*(d+1)}{E^*(d)} \tag{5}$$

where

$$E^*(d) = \frac{1}{N-d\tau} \sum_{i=1}^{N-d\tau} \left| x_{i+d\tau} - x_{n(i,d)+d\tau} \right| \tag{6}$$

2.2 Least square regression trees

The CART [16] involves classification tree and regression tree. The classification tree deals with a qualitative output variable whilst the regression tree handles a quantitative one. Given a data set comprised n couples of observation $(y_1, \boldsymbol{x}_1), \ldots, (y_n, \boldsymbol{x}_n)$, where $\boldsymbol{x}_i = (x_{1_i}, \ldots, x_{d_i})$ is a set of independent variables and $y_i \in R$ is a response associated with $\boldsymbol{x}_i$, the regression tree is constructed by using recursively partitioning process of this data set into two descendant subsets which are as homogeneous as possible until the terminal nodes are achieved.

The split values for partitioning process are chosen so that the sums of square errors are minimum. The sum of square error of the tth subset is expressed as:

$$R(t) = \frac{1}{n} \sum_{y_i, \boldsymbol{x}_i \in t} (y_i - \bar{y}(t))^2 \tag{7}$$

where $\bar{y}(t)$ and n are the mean value of response and the number of samples in that subset, respectively. At each terminal node, the predicted response is estimated by the average of all values of the response variables associated to that node. This issue reduces the accurate prediction significantly.

In the LSRT, the average of response at any node is replaced by the local model $f(\boldsymbol{\theta}, \boldsymbol{x}_i)$ which shows the relationship between the response y_i and a set of independent variable $\boldsymbol{x}_i$. Hence, the sum of square error of the tth node (subset) can be rewritten as:

$$R(t) = \frac{1}{n} \sum_{y_i, \boldsymbol{x}_i \in t} (y_i - f(\boldsymbol{\theta}, \boldsymbol{x}_i))^2 \quad (8)$$

where $\boldsymbol{\theta}$ is a set of parameters. The local models $f(\boldsymbol{\theta}, \boldsymbol{x}_i)$ can be either linear or non-linear model in which the forms are known with unknown values of parameters as shown in Table 1

Table 1 Local model types in LSRT

Model type	Description	Parameters
Polynomial	$y = \sum_{i=1}^{n+1} \theta_i x^{n+1-i}$	θ_i
Power	$y = \theta_1 x^{\theta_2}$ $y = \theta_1 + \theta_2 x^{\theta_3}$	$\theta_1, \theta_2, \theta_3$
Fourier	$y = \theta_0 + \sum_{i=1}^{n} \theta_{1_i} \cos(n\omega x) + \sum_{i=1}^{n} \theta_{2_i} \sin(n\omega x)$	$\theta_0, \theta_{1_i}, \theta_{2_i}$
Sine	$y = \sum_{i=1}^{n} \theta_{1_i} \sin(\theta_{2_i} x + \theta_{3_i})$	$\theta_{1_i}, \theta_{2_i}, \theta_{3_i}$

Those local models are organized as a set of models. At any node, the values of parameters of each model are initially calculated by using least square method [11], the fit model are subsequently chosen based on the sum of squares due to error (SSE) and the root mean squared error (RMSE) criterions:

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (9)$$

where y_i and $\hat{y}_i$ are response value and predicted value given by local model at that node, respectively. Consequently, the outputs of terminal nodes are local models that lead to more accurate prediction.

Similarly to CART, LSRT is also pruned in order to avoid the overfitting and complicated problems. In this work, we use 10 cross-validations to select the best tree size.

3. Proposed system

Normally when a fault occurs, the conditions of machine can be identified by the change in vibration amplitude. In order to predict the future state based on available vibration data, the proposed system as shown in Fig. 2 which consists of four procedures is proposed.

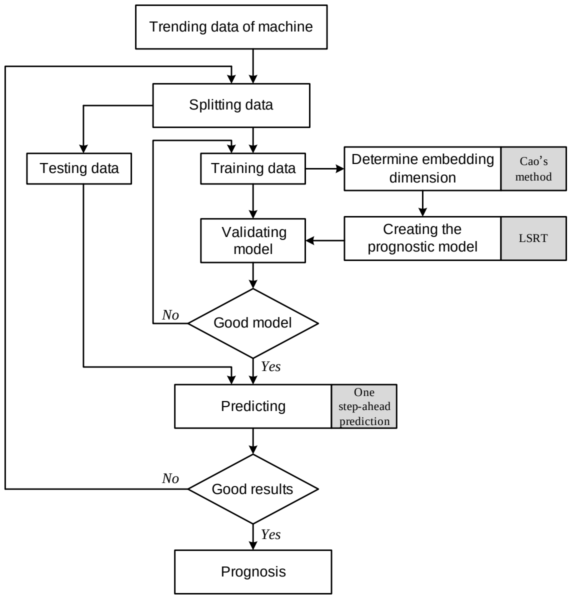

Fig. 2 Proposed system for machine fault prognosis.

The role of each procedure is explained as follows:

Step 1 Data acquisition: acquiring vibration signal during the running process of the machine until faults occur.

Step 2 Data splitting: the trending data is split into two parts: training data for building the model and testing data for testing the validated model.

Step 3 Training-validating: determining the embedding dimension based on Cao's method, building the model and validating the model for measuring the performance capability.

Step 4 Predicting: one-step-ahead prediction is used to forecast the future value. The predicted result is measured by the error between predicted value and actual value in the testing data. If the prediction is successful, the result obtained from this procedure is the prognosis system.

4. Experiments and results

The proposed method is applied to real system to predict the trending data of a low methane compressor. This compressor shown in Fig. 3 and its specification is summarized in Table 2.

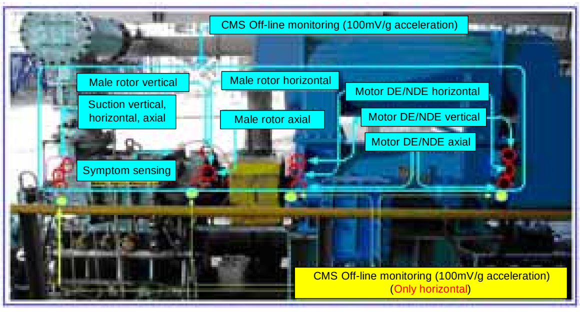

Fig. 3 Low methane compressor.

Table 2 Description of system

Electric motor		Compressor	
Voltage	6600 V	Type	Wet screw
Power	440 kW	Lobe	Male rotor(4 lobes)
Pole	2 Pole		Female rotor (6 lobes)
Bearing	NDE:#6216 DE:#6216	Bearing	Thrust: 7321 BDB
RPM	3565 rpm		Radial: Sleeve type

The data applied in this study is peak acceleration and envelope acceleration trending data recorded from August 2005 to November 2005 as shown in Figs. 4 and 5. Consequently, it can be seen as time-series data.

The machine is in normal condition during the first 300 points. After that time, the condition of machine suddenly changes indicating some faults occurring in this machine. With the aim of forecasting the change of machine condition, the first 300 points were used to train the system and the following 200 points were employed for testing system.

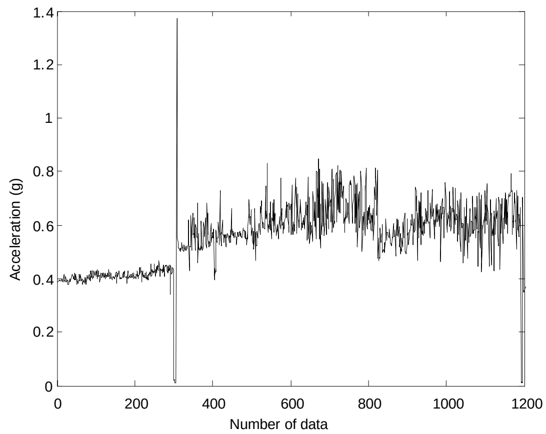

Fig. 4 The entire of peak acceleration data of low methane compressor.

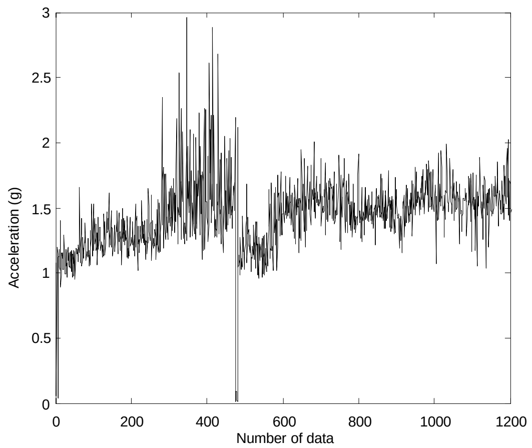

Fig. 5 The entire of envelope acceleration data of low methane compressor.

The predicting performance is evaluated by using the RMSE given in Eq. (9). The time delay value is chosen as 1 for the reason that one step-ahead is implemented in all datasets. The embedding dimension is estimated to be 6 when the values of $E_1(d)$ reaches its saturation as depicted in Fig. 6.

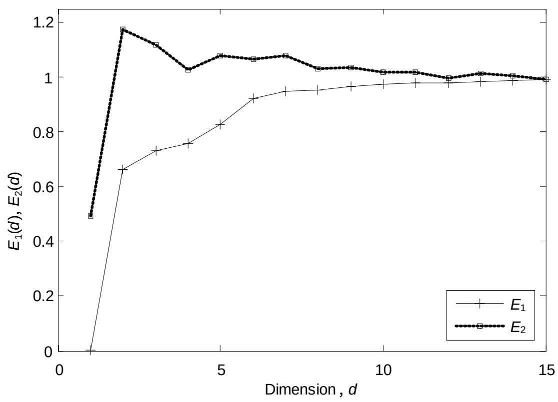

Fig. 6 The values of E_1 and E_2 of peak acceleration data of low methane compressor.

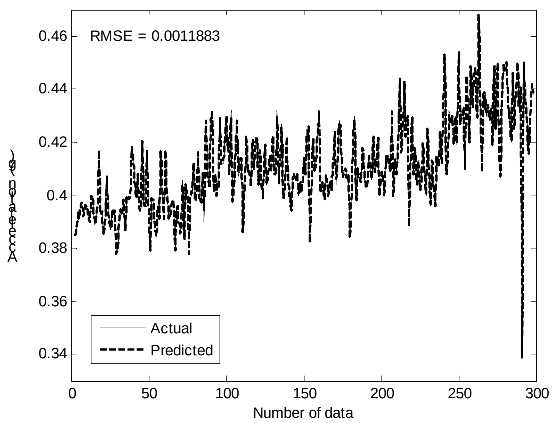

Fig. 7 Training and validating results of peak acceleration data.

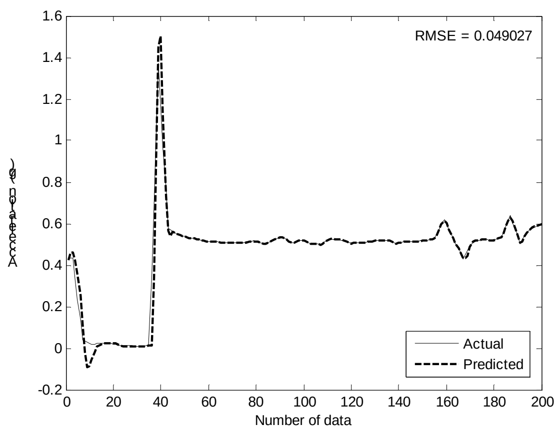

Fig. 8 Predicted results of peak acceleration data using LSRT.

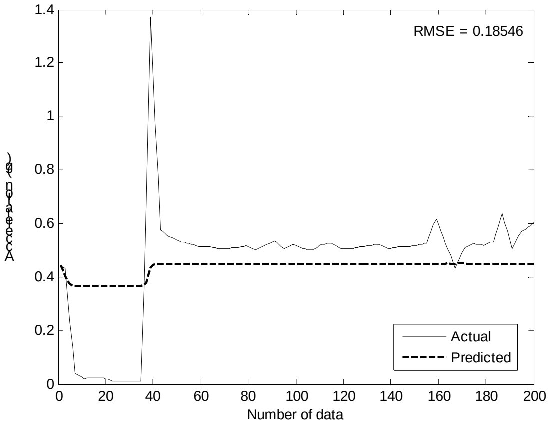

Fig. 9 Predicted results of peak acceleration data using CART.

Fig. 7 depicts the training and validating results of peak acceleration data with a small RMSE value of 0.00118. In testing process, the independent data set contained the changing machine condition is used. Fig.8 shows the actual-like predicted results with the RMSE error of 0.049027 although the predicting model was not trained with those changing values. That is impossible to obtain with CART as shown in Fig. 9.

Table 3 The RMSE of CART and LSRT

Data type	Training		Testing	
	CART	LSRT	CART	LSRT
Peak acceleration	0.00062	0.0011	0.1855	0.049
Envelop acceleration	0.00028	0.00015	0.1429	0.101

Table 3 shows not only the remaining results of applying LSRT on envelop acceleration data but also the comparison of the RSME between CART and LSRT. According to table 3, training results of CART are sometimes slightly smaller than those of LSRT but the testing results of CART are always larger. This shows the superior of LSRT in aspect of machine condition prognosis.

5. Conclusions

Machine condition prognosis is extremely significant in foretelling the degradation of working condition and trends of fault propagation before they reach the alarm. In this study, the least square regression tree together with one-step-ahead of time-series techniques have been investigated for machine condition prognosis. The proposed method is validated by predicting future state condition of a low methane compressor wherein the peak acceleration and envelope acceleration have been examined. The obtained results confirm that the proposed method offers a potential for machine condition prognosis with one-step-ahead prediction.

References

[1] Jianhui Luo, M. Namburu, K. Pattipati, Liu Qiao, M. Kawamoto, S. Chigusa, Model-based prognostic techniques, AUTOTESTCON Proceedings of IEEE Systems Readiness Technology Conference, 22-25 Sept. (2003) 330–340.

[2] M. J. Roemer, C. S. Byington, G. J. Kacprzynski, G. Vachtsevanos, An overview of selected prognostic technologies with application to engine health management, Proceedings of GT2006, (2006).

[3] G. Vachtsevanos, P. Wang, Fault prognosis using dynamic wavelet neural networks, AUTOTESTCON Proceedings, IEEE Systems Readiness Technology Conference, 22-23 Aug. (2001) 857–870.

[4] R. Huang, L. Xi, X. Li, C. R. Liu, H. Qiu, J. Lee, Residual life prediction for ball bearings based on self-organizing map and back propagation neural network methods, Mechanical Systems and Signal Processing 21 (2007) 193–207.

[5] W.Q. Wang, M.F. Golnaraghi, F. Ismail, Prognosis of machine health condition using neuro-fuzzy system, Mechanical System and Signal Processing 18 (2004) 813–831.

[6] W. Wang, An adaptive predictor for dynamic system forecasting, Mechanical Systems and Signal Processing 21 (2007) 809–823

[7] M.B. Kennel, R. Brown, H.D.I. Abarbanel, Determining embedding dimension for phase-space reconstruction using a geometrical construction, Physical Review A 45 (1992) 3403–3411.

[8] L. Cao, Practical method for determining the minimum embedding dimension of a scalar time series, Physica D 110 (1997) 43–50.

[9] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and regression trees, Chapman & Hall (1984).

[10] J. Yang, J. Stenzel, Short-term load forecasting with increment regression tree, Electric Power Systems Research 76 (2006) 880–888

[11] R. Johansson, System modeling and identification, Prentice-Hall International, Inc., (1993).

[12] A. Suárez, J.F. Lutsko, Globally optimal fuzzy decision trees for classification and regression, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (1999) 1297-1311.

(13) C. Huang, J.R.G. Townshend, A stepwise regression tree for nonlinear approximation: application to estimating subpixel land cover, International Journal of Remote Sensing 24 (2003) 75-90.

[14] D.S. Vogel, O. Asparouhov, T. Scheffer, Scalable look-ahead linear regression trees, Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2007).