Machine condition prognosis based on regression trees and one-step-ahead prediction

Van Tung Tran[a], Bo-Suk Yang[a,*], Myung-Suck Oh[a], Andy Chit Chiow Tan[b]

[a]*School of Mechanical Engineering, Pukyong National University, San 100, Yongdang-dong, Namgu, Busan 608-739, South Korea*

[b]*School of Mechanical, Manufacturing and Medical Engineering, Queensland University of Technology, G.P.O. Box 2343, Brisbane, Qld. 4001, Australia*

Abstract

Predicting the degradation of working conditions of machinery and trending of fault propagation before they reach the alarm or failure threshold is extremely important in industry to fully utilize the machine production capacity. This paper proposes a method to predict the future conditions of machines based on one-step-ahead prediction of time-series forecasting techniques and regression trees. In this study, the embedding dimension is firstly estimated in order to determine the necessarily available observations for predicting the next value in the future. This value is subsequently utilized for the predictor which is generated by using regression tree technique. Real trending data of low methane compressor acquired from condition monitoring routine are employed for evaluating the proposed method. The results indicate that the proposed method offers a potential for machine condition prognosis.

Keywords: Embedding dimension; Regression trees; Prognosis; Time-series forecasting

1. Introduction

Unexpected catastrophic failures of machine that lead to a costly maintenance or even human casualties can be avoided with the proviso that the machine is appropriately maintained. Traditional maintenance strategies commonly used in industry consist of corrective maintenance and preventive maintenance. The former means “fix it when it breaks”, i.e. maintenance is carried out after a breakdown or when an obvious fault has occurred, whilst the latter is carried out in order to prevent equipment breakdown by performing repair, service, or replacing components at a fixed schedule. Even though preventive maintenance plans increase the

reliability of machine, they are costly due to the frequent replacements of the expensive components before the end of their lives and the reduction of the availability of the machine's productive capability. Therefore, the strategies of traditional maintenance are not adequate to fulfill the needs of expensive and high availability of industrial systems.

Condition-based maintenance (CBM) which involves prognostic module is an alternative maintenance strategy that allows the machine to operate continuously until symptoms of a failure are detected. Prognosis is the ability to access the current state, forecast the future state, and predict the time-to-failure or the remaining useful life (RUL) of a failing components or subsystems. The RUL is the time left for the normal operation of machine before the breakdown occurs or machine condition reaches the critical failure value. Prognosis is also used to produce warning when the machine condition reaches the predetermined setup alarm or critical failure threshold. Furthermore, it can be used for running repairs periodically in manufacturing facilities and fault-tolerant control [1]. Due to the benefits mentioned above, prognosis has been extensively researched with focus on condition-based maintenance in the recent time.

Nevertheless, prognosis is a relatively new area and becomes a significant part of CBM of systems [2]. Currently, numerous approaches to prognosis that range in reliability from simple historical failure rate models to physics-based models have been developed. According to [3], the hierarchy of potential prognostic approaches related to their applicability and relative accuracy as well as their complexity is performed in Fig. 1. Out of the experience-based prognostic technique that requires the component failure history data, the remaining techniques use the models for the purposes of predicting the future conditions of the monitored system.

Fig. 1 Hierarchy of prognostic approaches

Model-based prognosis techniques require an accurate mathematical model of the failure modes to be used to predict the RUL of critical components. Some of the published researches in prognosis include predicting RUL of high power clutch systems, estimating the time-to-failure of mobile robot and forecasting the remaining utility of bearings [4-6]. However, those techniques are merely applied for some specific components and each of them needs a different mathematical model. Furthermore, a suitable model is also difficult to establish to mimic the real life.

The data-driven approaches are directly derived from the routinely monitored system operating data and associated with either statistical or learning techniques based on the theory of pattern recognition. Artificial intelligent techniques which belong to learning techniques are regularly considered by researchers due to their flexibility in generating appropriate models.

The salient researches based on data-driven approaches have been proposed, such as Vachsevanos and Wang [7] who used dynamic wavelet neural network to predict the failure growth based on the vibration signals to estimate the RUL of bearings. Huang et al. [8] predicted the RUL of ball bearing by self-organizing map and back propagation neural networks methods using vibration signals. Wang et al. [9] utilized and compared the results of two predictors, which were recurrent neural networks and adaptive neuro-fuzzy inference systems, to forecast the damage propagation trend of rotating machinery. A hybrid approach of fuzzy logic and neural networks was employed to predict the RUL of bearings of small and medium size induction motors [10].

Time series prediction is a problem encountered in many fields from engineering (predictive control of industrial processes) to finance (forecasting returns of shares or stock markets). Models and prediction methodologies have been proposed by a large community of researchers. For examples, Thissen et al. [11] applied three methods, namely, support vector machines, recurrent neural networks and autoregressive moving average for times-series which predicted the process chemo metrics in field. Dulakshi et al. [12] employed artificial neural networks on time-series to predict river flow. Simon [13] used self-organizing maps algorithm to predict the missing value in CATS benchmark dataset.

The problem to be dealt with when predicting the future value is how many steps (time delays) are appropriate for obtaining the best performance? In time-series forecasting techniques, one-step-ahead or multi-step-ahead prediction is frequently used. One-step-ahead or multi-step-ahead prediction implies that the predictor utilizes the available observations to forecast one value or multiple values at the definite future time. According to Wang [1], the more the steps ahead is, the less reliable the forecasting operation is because multi-step prediction is associated with multiple one-step operations. This issue will be discussed further in the experiments and results section.

Another problem is how many essential observations (inputs) are used for forecasting the future value, so-called embedding dimension d. The value of d should be chosen large enough for the predictor to estimate accurately the future value of machine condition and not too large to avoid the unnecessary increase in computational complexity. False nearest neighbor method (FNN) [14] and Cao's method [15] are commonly used to determine the embedding dimension. However, FNN method depends on the chosen parameters wherein different values lead to different results. Furthermore, FNN method also depends on the number of available observations and is sensitive to additional noise. Cao's method overcomes the shortcomings of the FNN approach and therefore, it is chosen in this study. After determining the embedding dimension d, the predictor is considered subsequently. Classification and regression tree (CART) [16] is widely implemented in machine fault diagnosis. In the prediction techniques,

regression tree and its extensions are applied to forecast the short-term load of the power system [17, 18] with excellent performance. Hence, regression tree is proposed as a predictor for machine prognosis in this paper.

2. Background knowledge

2.1. Determine the embedding dimension

Assuming a time-series of x_1, x_2, …, x_N. The time delay vector constructed from this time series is defined as follows:

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

where τ is the time delay and d is the embedding dimension. 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 ‖·‖ 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 so 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, a 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 only dependent on 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 increasing when the time series comes from a deterministic process. If a plateau is observed for $d \geq d_0$ then $d_0 + 1$ is the minimum embedding dimension.

The Cao's method also introduced another quantity $E_2(d)$ to overcome the problem in practical computations where $E_1(d)$ is slowly increasing or has stopped changing if d is sufficiently large:

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

where

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

According to [15], for purely random process, $E_2(d)$ is independent of d and equal to 1 for any of d. However, for deterministic time-series, $E_2(d)$ is related to d. Consequently, there must exist some d's so that $E_2(d) \neq 1$.

2.2. Regression trees

CART method has been extensively developed by Breiman et al. [16] for classification or regression purpose depending on the response variable which is either categorical or numerical. In this study, CART is utilized to build up a regression tree model. Beginning with an entire data set, a binary tree is constructed with the repeated splits of the subsets into two descendant subsets according to independent variables. The goal is to produce subsets of the data which are as homogeneous as possible with respect to the response variables. Regression tree in CART is built by using the following two processes: tree growing and tree pruning.

2.2.1. Tree growing

Let L be a learning sample of size n, and it comprises n couples of observations $(y_1, \boldsymbol{x}_1),...,(y_n, \boldsymbol{x}_n)$ where $\boldsymbol{x}_i = (x_{1_i},...,x_{d_i})$ is a set of independent variables and $y_i \in R$ is a response associated with $\boldsymbol{x}_i$. The objective of regression tree is to predict the values of response variables $y = (y_1,..., y_n)$ derived from the set of independent variables $(\boldsymbol{x}_1,..., \boldsymbol{x}_n)$. In order to build the tree, learning sample L is partitioned into two subsets by binary split. Splits are formed by using the inequality condition between criterion and value of independent variables. The result of this splitting is to move the couples $(y, \boldsymbol{x})$ to left or right nodes containing more homogeneous responses. This process is repeated until the terminal nodes are achieved. The predicted response at each terminal node t is estimated by the average $\bar{y}(t)$ of the $n(t)$ values y contained in that terminal node. The final structure of a binary tree T is shown in Fig. 2.

Fig. 2. Binary regression tree T.

The split selection at any internal node t is chosen according to the node impurity that is measured by within-node sum of squares. The within-node sum of squares is the squared difference between the predicted response and observed response:

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

and, $\bar{y}(t) = \frac{1}{n(t)} \sum_{y_i, \boldsymbol{x}_i \in t} y_i$. (8)

When a split is performed, two subsets of observations t_L and t_R are obtained. The optimum split s^* at node t is obtained from the set of all splitting candidates S in order that it qualifies:

$$\begin{aligned} &\Delta R(s^*, t) = \max \Delta R(s, t); \quad s \in S \\ &\Delta R(s, t) = R(t) - R(t_L) - R(t_R) \end{aligned} \tag{9}$$

where $R(t_L)$ and $R(t_R)$ are sum of squares of the left and right subsets, respectively.

The tree gained in tree growing process has many terminal nodes that lead to increase precision in estimating of the responses. However, the tree with many terminal nodes is frequently too complicated and overfitting is highly probable. Consequently, it should be pruned back to select the best tree.

2.2.2. Tree pruning

Let $T_{\max}$ denote tree built on the sample L and T denote a subtree attained from $T_{\max}$ by successively removing the nodes. Pruning a tree allows to decrease the size of $T_{\max}$, therefore T is smaller than $T_{\max}$. The result of pruning process will be a decreasing sequence of trees:

$$T_{\max} > T_1 > \ldots > T_k = \{t_1\} \tag{10}$$

where $\{t_1\}$ denotes the root of tree. Along with the sequence of pruned trees, a corresponding sequence of value α is found:

$$0 = \alpha_1 < \alpha_2 < \ldots < \alpha_k < \alpha_{k+1} < \ldots < \alpha_K \tag{11}$$

where $\alpha \geq 0$ is the cost parameter which weights the number of terminal node. For $k \geq 1$, the tree T_k is the smallest subtree that minimizes the error-complexity for interval $\alpha_k < \alpha < \alpha_{k+1}$, and $T(\alpha) = T(\alpha_k) = T_k$.

The error-complexity is defined as:

$$R_\alpha(T) = R(T) + \alpha \left| \tilde{T} \right| \tag{12}$$

where $R(T) = \frac{1}{n} \sum_{t \in \tilde{T}} \sum_{(y_i, \boldsymbol{x}_i) \in t} \left(y_i - \bar{y}(t) \right)^2$ is the total within-node sum of squares, $\tilde{T}$ is the set of current nodes of T and $\left| \tilde{T} \right|$ is the number of terminal nodes in T.

Tree pruning process is performed by the following procedure:

Step 1: At every internal node, an error-estimation is found for the number of descendant subtrees.

Step 2: Using the error-estimation attained at step 1, the internal node with the smallest error is replaced by terminal node.

Step 3: The algorithm terminates if all the internal nodes have converged to a terminal node. Otherwise, it returns to step 1.

The best tree is chosen when both the error-complexity and the error-estimation are minimized. The minimum error-estimation could be obtained by using the largest tree, but this increases the error-complexity. Therefore, the trade-off between those two criterions should be considered. There are two possible methods to accomplish this. One is through the use of independent test samples and the other is cross-validation. In this study, cross-validation method is used to determine the error-estimation.

2.2.3. Cross-validation for selecting the best tree

The learning sample L is randomly divided into V mutually exclusive data sets $L_1, \ldots, L_v$. It is the best to ensure that all the V learning samples are of the same size or nearly the same. Let the learning sample $L^{(v)}$ be the vth learning sample represented by $L^{(v)} = L - L_v$, $v = 1,...,V$. $L^{(v)}$ is used as a learning sample for model building and L_v is reserved as a testing sample for determining the error-estimation of the model. Each learning sample is used to grow a large tree and to get the corresponding sequence of pruned subtrees. Thus, a sequence of trees $T^{(v)}(\alpha)$ that represents the minimum error-complexity trees for given values of α is achieved.

Simultaneously, the entire learning sample L is also employed to grow a large tree and to get the sequence of subtrees T_k with the corresponding sequence of α_k. Define $\alpha'_k = \sqrt{\alpha_k \alpha_{k+1}}$ as the geometric midpoint of the interval $[\alpha_k, \alpha_{k+1}]$. Use $d_k^{(v)}(\boldsymbol{x})$ to denote the prediction corresponding to the tree $T^{(v)}(\alpha'_k)$. The cross-validation which estimates for the error-estimation is given:

$$R^{CV}\left(T_k(\alpha'_k)\right) = \frac{1}{n}\sum_{v=1}^{V}\sum_{(y_i,\boldsymbol{x}_i)\in L_v}\left(y_i - d_k^{(v)}(\boldsymbol{x}_i)\right)^2 \tag{13}$$

From each case of the testing sample L_v with $d_k^{(v)}(\boldsymbol{x})$, a predicted response is attained, and then the squared difference is calculated between the predicted response and actual response. This process is repeated for every testing sample. The average value of these errors is taken to determine the error-estimation of a tree based on values of R^{CV}.

The standard error of cross-validation is obtained by using

$$S\hat{E}\left(\hat{R}^{CV}(T_k)\right) = \sqrt{\frac{s^2}{n}} \tag{14}$$

and

$$s^2 = \frac{1}{n}\sum_{\boldsymbol{x}_i, y_i}\left[\left(y_i - d_k^{(v)}(\boldsymbol{x}_i)\right)^2 - \hat{R}^{CV}(T_k)\right]^2 \tag{15}$$

The subtree that has the smallest error is found and denoted by T_0. Finally, the best tree T_k^*

is selected such that

$$\hat{R}^{CV}(T_k^*) \le \hat{R}^{CV}_{\min}(T_0) + S\hat{E}\left(\hat{R}^{CV}_{\min}(T_0)\right)$$
$$\hat{R}^{CV}_{\min}(T_0) = \min\left\{\hat{R}^{CV}(T_k)\right\} \tag{16}$$

3. Proposed system

Normally, when a fault occurs in a machine, the conditions of machine can be identified by the change in vibration amplitude. In order to predict the future state of the machine conditions based on available vibration data, regression tree predictor is used. The proposed system is shown in Fig. 3. This system consists of four procedures sequentially: data acquisition, data splitting, training-validating model and predicting. The role of each procedure is explained as follows:

Step 1 Data acquisition: the use of vibration data from the machine for trending. It consists of historical data and vibration data during the running process of the machine until faults occur.

Step 2 Data splitting: the trending data is split into two parts: training data and testing data. Different data is used for different purposes in the prognosis system. Training data is used for building the model whilst testing data is utilized to test the validated model.

Step 3 Training-validating: this procedure includes the following sub-procedures: determining the embedding dimension based on Cao's method, building the model and validating the model. Validating model is used for measuring the performance capability of the model.

Step 4 Predicting: this procedure uses one-step-ahead prediction 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.

Fig. 3. Proposed system for machine fault prognosis.

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. 4 is driven by a 440 kW motor, 6600 volt, 2 pole and operating at speed of 3565 rpm. The other information of system is summarized in Table 1.

Fig. 4. Low methane compressor: wet screw type.

The data applied in this study is peak acceleration and envelope acceleration trending data. The trending data was recorded from August 2005 to November 2005. The average sampling rate was 6 hours during the data acquisition process. This data approximately consists of 1200 data points shown in Figs. 5 and 6. This data contains information of machine history with respect to time sequence (vibration amplitude). Consequently, it can be seen as time-series data. The proposed method is employed to predict the future condition of vibration amplitude based on the past and current states.

Table 1 Information of system

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

Fig. 6. The entire of envelope acceleration data of low methane compressor.

The machine is in normal condition during the time correlated with the first 300 points. After that time, the condition of machine suddenly changes. It indicates that there are some faults occurring in this machine. These faults were the damages of main bearings of compressor (notation Thrust: 7321 BDB) due to the insufficient lubrication. Consequently, the babbit surfaces of these bearings were overheated and delaminated as depicted in Fig.7.

Fig. 7. The faults of main bearings of compressor.

With the aim of forecasting the change of machine condition, the first 300 points were used to train the system and the following 150 points were employed for testing system. In order to evaluate the predicting performance, the root-mean square error (RMSE) is utilized as following

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}} \tag{17}$$

where N represents the total number of data points in the test set, y_i is actual value in training set or test set and $\hat{y}_i$ represents the predicted value of the model.

Because the sampling rate was 6 hours in data acquisition process, it is large enough to make the decision whether to stop or to continue the operation of this machine when its conditions reach the predetermined setup alarm or critical failure threshold. That is the reason for choosing one-step-ahead prediction in this study. Therefore, the time delay value is also chosen as $\tau = 1$ in all datasets. Furthermore, the number of cases for each terminal node in tree growing process is 5 and 10 cross-validations are decided for selecting the best tree in tree pruning.

Fig. 8. Training and validating results of peak acceleration data (the first 300 points).

Fig. 9. Predicted results of peak acceleration data.

The model is built and the best tree size is determined by using 10 cross-validations. Excellent validating performance is shown in Fig. 8 with a small RMSE value of 0.00062. However, in the testing process, an unexpected result occurs as depicted in Fig. 9. It shows that the model is incapable of predicting the future value. The reason is that the model could be improperly trained. In other words, the training data does not contain anomalous values that lead to the incapability of predicting accurately. This affirmation could be demonstrated by using another data set consisting of those anomalous values as shown in Fig. 10.

Theoretically, the minimum embedding dimension is chosen as $E_1(d)$ obtains a plateau. In Fig.11, the embedding dimension is chosen as 6 for the reason that the values of $E_1(d)$ reaches its saturation. This value is used for regression tree predictor. Figs. 12 and 13 are the validating and testing model, respectively. The training and validating results of peak acceleration data are almost identical, as shown in Fig. 12, with a very small RMSE value of 0.000601. In the testing process, even though the model cannot predict accurately the machine condition, the RMSE value shown in Fig. 13 is 0.0143 which is acceptable.

Fig. 10. Peak acceleration of low methane compressor.

Fig. 11. The values of E_1 and E_2 of peak acceleration data of low methane compressor.

Fig. 12. Training and validating results of peak acceleration data.

Fig. 13. Predicted results of peak acceleration data.

Fig. 14 performs the trending data of envelope acceleration of low methane compressor. By using the similar processes i.e. estimating embedding dimension (this value is 6 shown in Fig. 15), validating model and testing model by using independent data set are respectively carried out and the final results are performed in Figs. 16 and 17. The training and validating results closely resemble the actual data with a RMSE error of 0.000291, as shown in Fig. 16. Although the predictor is incapable of predicting the machine condition precisely, it can closely track the changes of trending condition of machine with a small error of 0.06 as shown in Fig. 17.

Fig. 14. Data trending of envelope acceleration of low methane compressor.

Fig. 15. The values of E_1 and E_2 of envelope acceleration data.

Fig. 17. Training and validating results of envelope acceleration data.

Fig. 18. Predicted results of envelope acceleration data.

5. Conclusions

Machine condition prognosis is extremely significant in foretelling the degradation working condition and trends of fault propagation before they reach the alarm or failure threshold. In this study, the machine prognosis based on one-step-ahead of time-series techniques and regression trees has been investigated. 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. Using 10 cross-validations to find the optimum tree size and an embedded dimension of 6, the results give a prediction error of 1.43% with peak acceleration data, and 6% with the enveloped acceleration data. These errors are small in statistical sense. The results confirm that the proposed method offers a potential for machine condition prognosis with one-step-ahead prediction.

References

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

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

[3] G. Vachtsevanos, F. Lewis, M. Roemer, A. Hess, B. Wu, Intelligent fault diagnosis and prognosis for engineering system, Wiley, 2006

[4] M. Watson, C. Byington, D. Edwards, S. Amin, Dynamic modeling and wear-based remaining useful life prediction of high power clutch systems, Tribology Transactions 48 (2005) 208–217.

[5] M. Luo. D. Wang, M. Pham, C.B. Low, J.B. Zhang, D.H. Zhang, Y.Z. Zhao, Model-based fault diagnosis/prognosis for wheeled mobile robots: a review, Industrial Electronics Society ,32nd Annual conference of IEEE, 6-10 Nov. (2005) 2267–2272.

[6] Y. Li, T.R. Kurfess, S.Y. Liang, Stochastic prognostics for rolling element bearings, Mechanical Systems and Signal Processing 14 (2000) 747–762.

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

[8] 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.

[9] 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.

[10] B. Satish, N.D.R. Sarma, A fuzzy BP approach for diagnosis and prognosis of bearing faults in induction motors, IEEE Power Engineering Society General Meeting 3 (2005) 2291–2294.

[11] U. Thissen, R. van Brakel, A.P. de Weijer, W.J. Melssen, L.M.C. Buydens, Using support vector machines for time series predicting, Chemometrics and Intelligent Laboratory Systems 69 (2003) 35–49.
[12] Dulakshi S.K. Karunasinghea, Shie-Yui Liongb, Chaotic time series prediction with a global model: Artificial neural network, Journal of Hydrology 323 (2006) 92–105.
[13] G. Simon, J.A. Lee, M. Cottrell, M. Verleysen, Forecasting the CATS benchmark with the Double Vector Quantization, Neurocomputing, in press.
[14] 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.
[15] L. Cao, Practical method for determining the minimum embedding dimension of a scalar time series, Physica D 110 (1997) 43–50.
[16] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and regression trees, Chapman & Hall (1984).
[17] J. Yang, J. Stenzel, Short-term load forecasting with increment regression tree, Electric Power Systems Research 76 (2006) 880–888.
[18] H. Moti, N. Kosemura, K. Ishiguro, T. Kondo, Short-term load forecasting with fuzzy regression tree in power systems, System Man and Cybernetics, IEEE International Conference 3 (2001) 1948–1953.

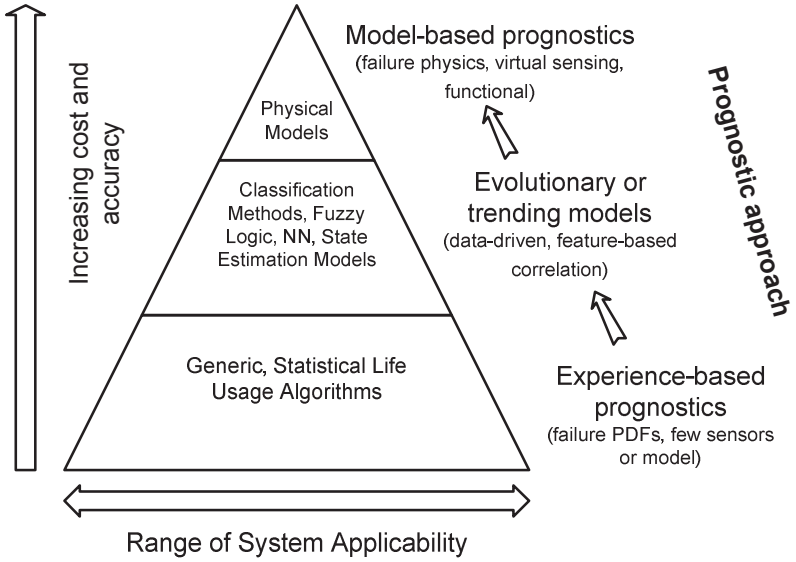

Fig. 1 Hierarchy of prognostic approaches

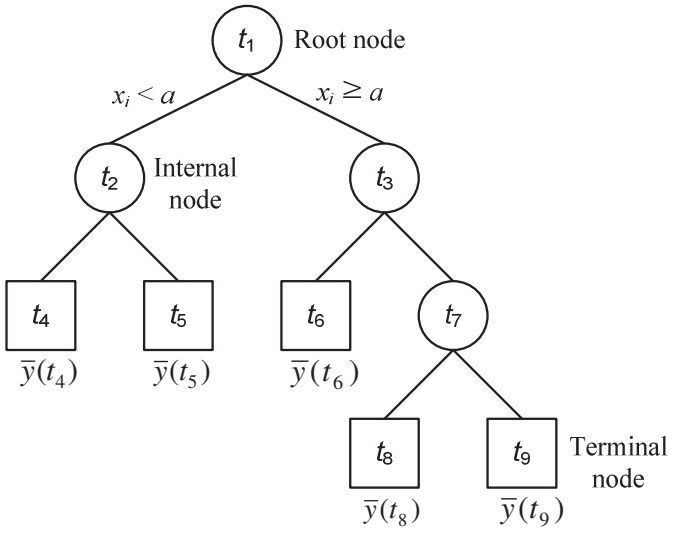

Fig. 2. Binary regression tree.

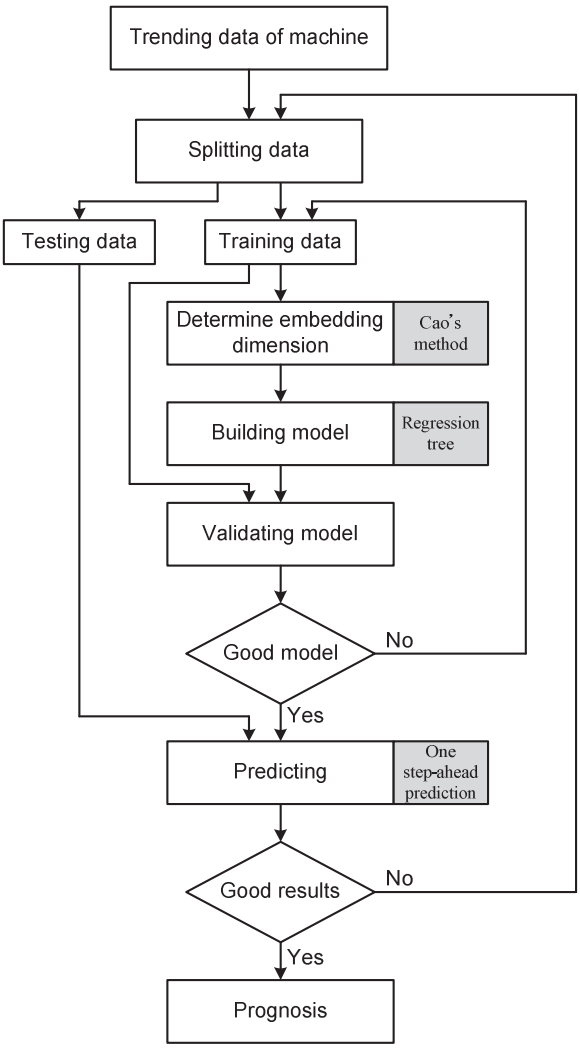

Fig. 3. Proposed system for machine fault prognosis.

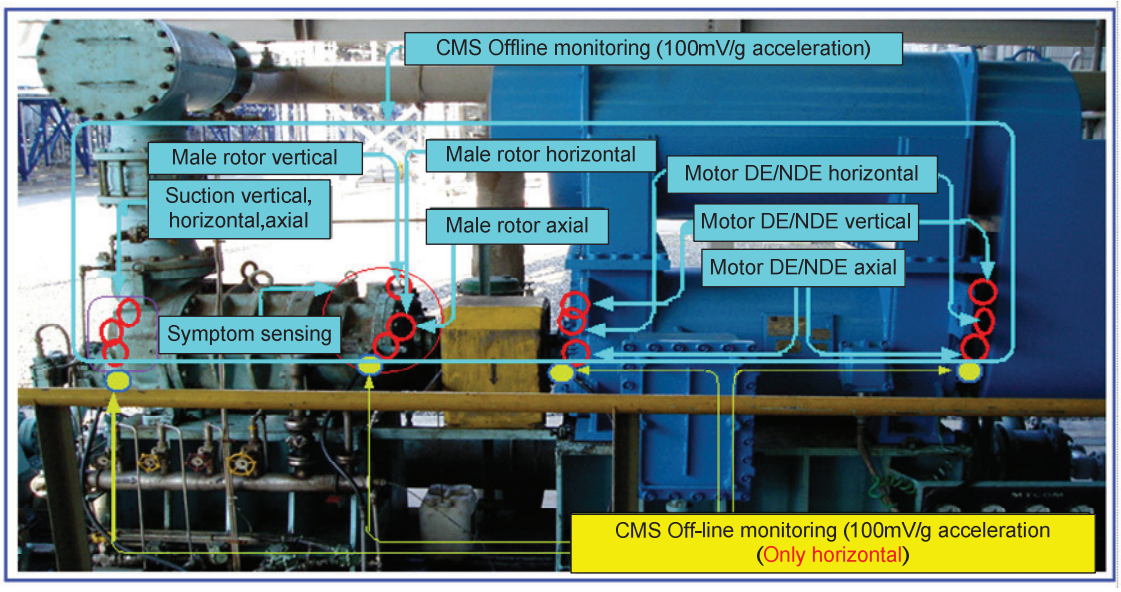

Fig. 4. Low methane compressor: wet screw type.

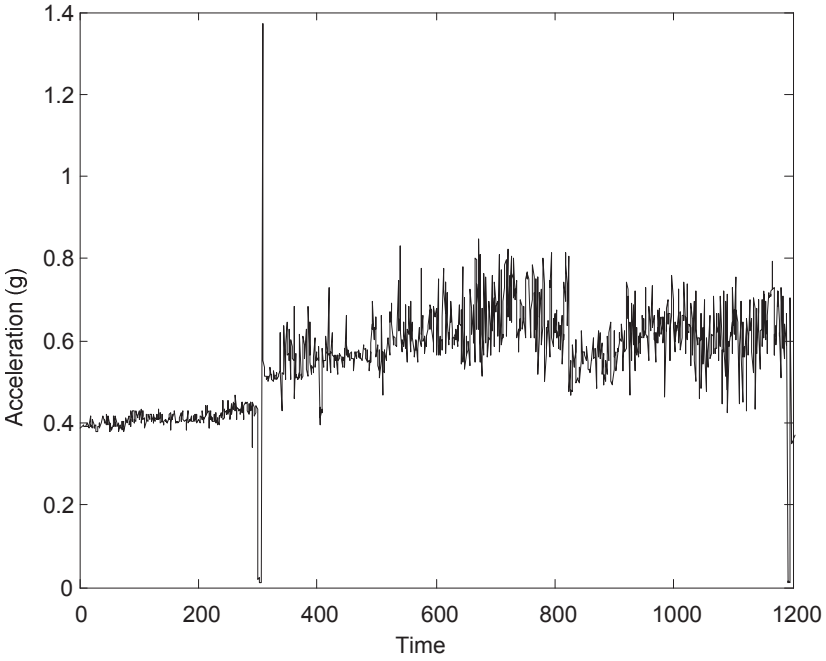

Fig. 5. The entire of peak acceleration data of low methane compressor.

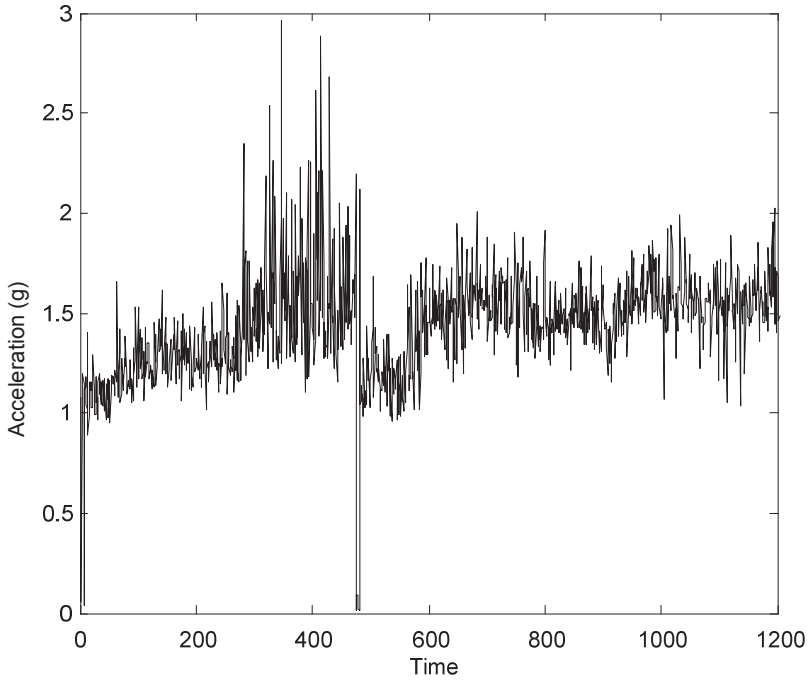

Fig. 6. The entire of envelope acceleration data of low methane compressor.

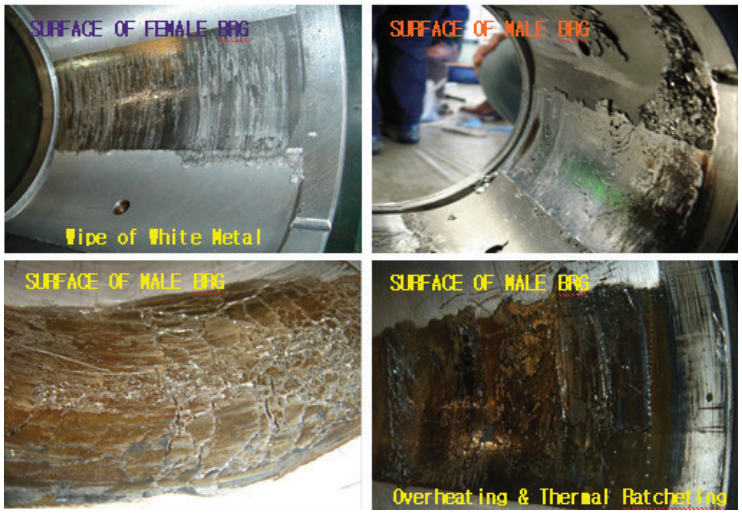

Fig. 7. The faults of main bearings of compressor.

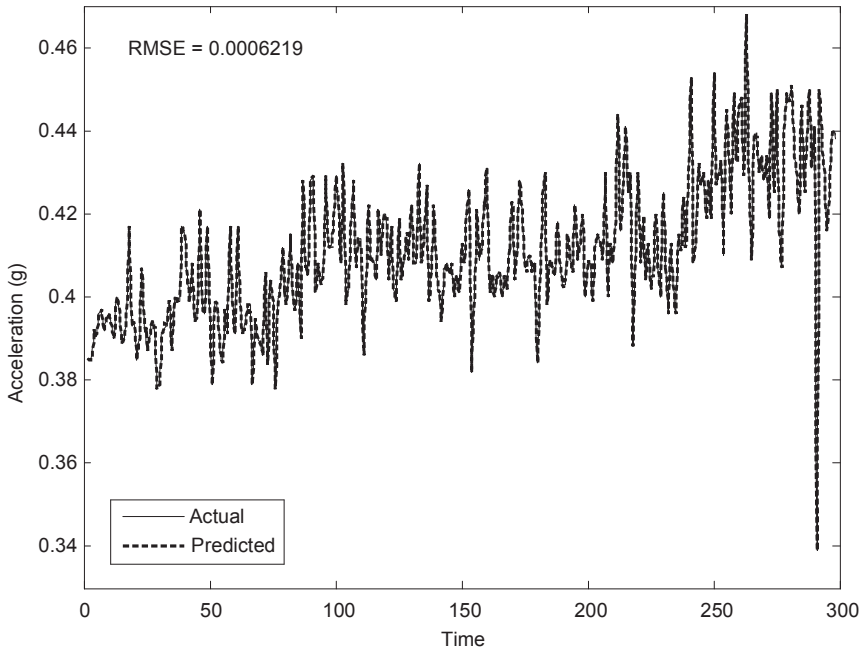

Fig. 8. Training and validating results of peak acceleration data (the first 300 points).

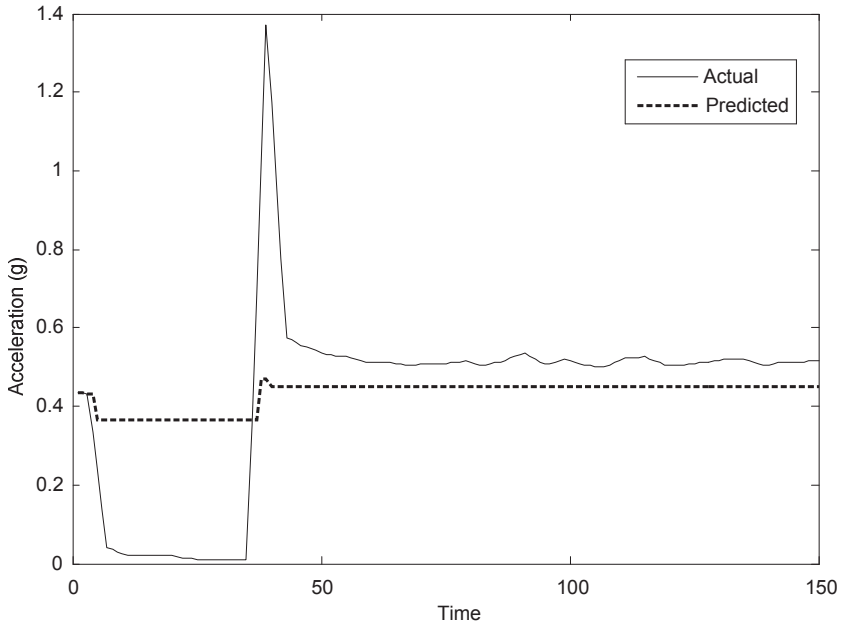

Fig. 9. Predicted results of peak acceleration data.

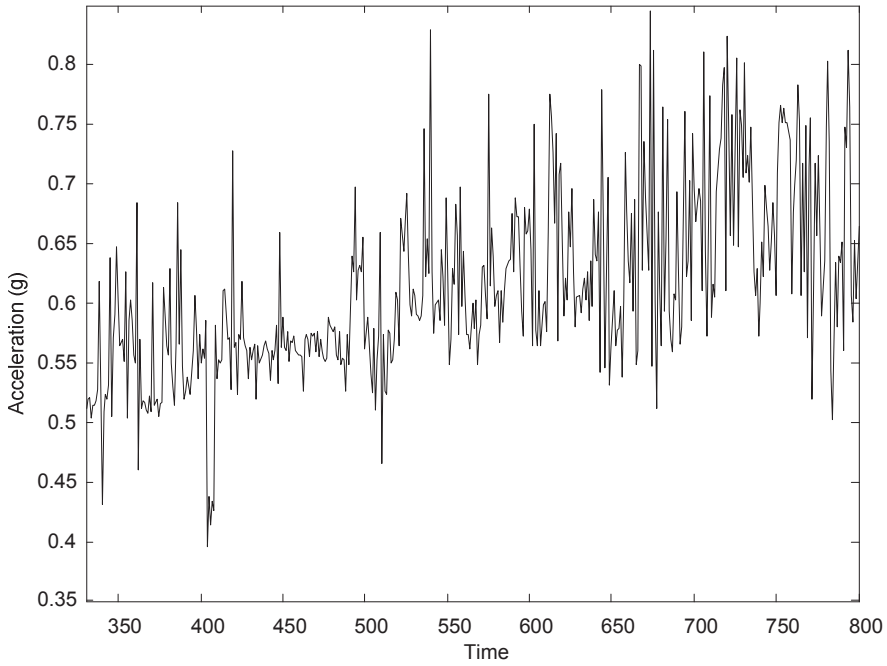

Fig. 10. Peak acceleration of low methane compressor.

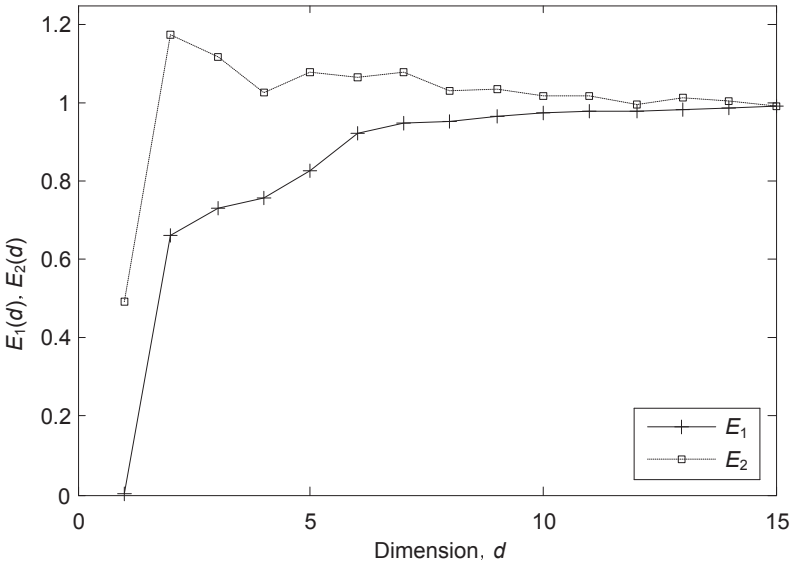

Fig. 11. The values of E_1 and E_2 of peak acceleration data of low methane compressor.

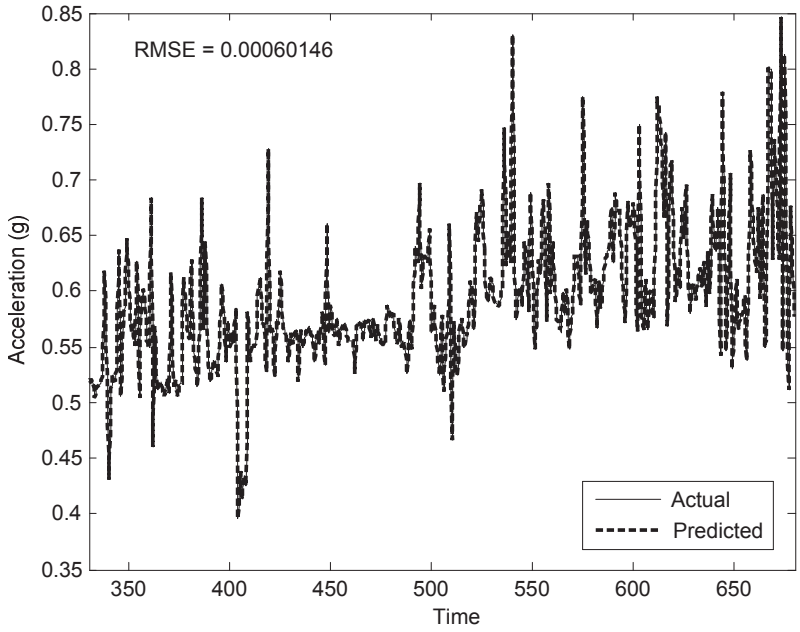

Fig. 12. Training and validating results of peak acceleration data.

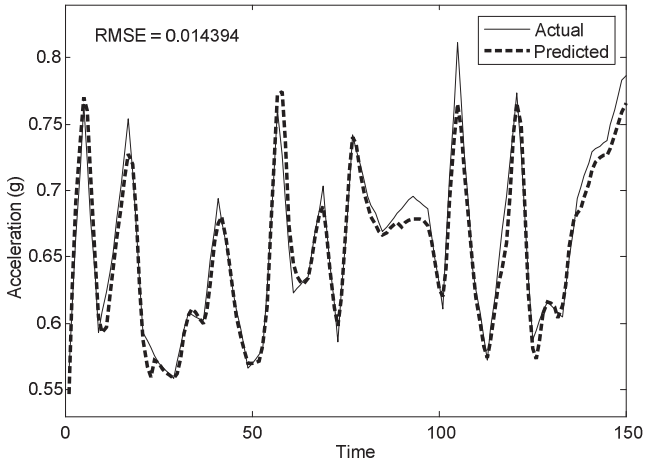

Fig. 13. Predicted results of peak acceleration data.

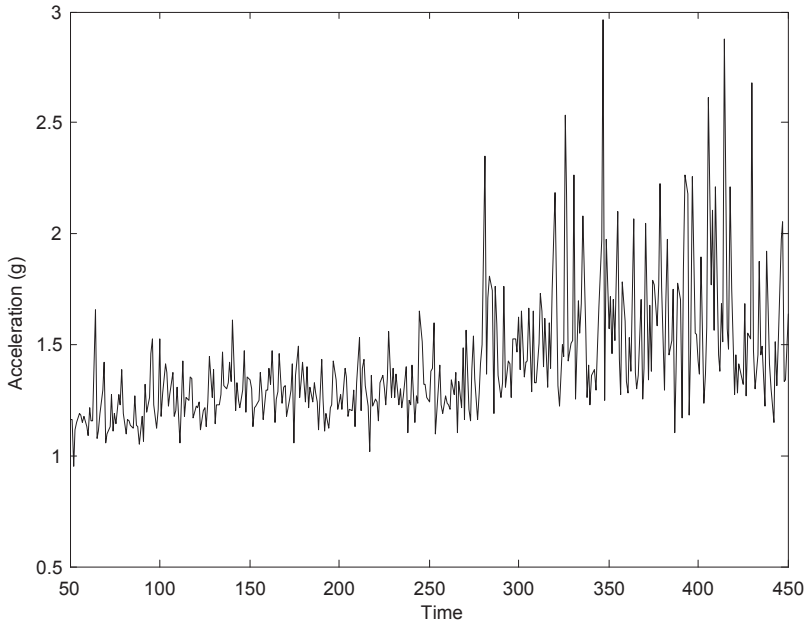

Fig. 14. Data trending of envelope acceleration of low methane compressor.

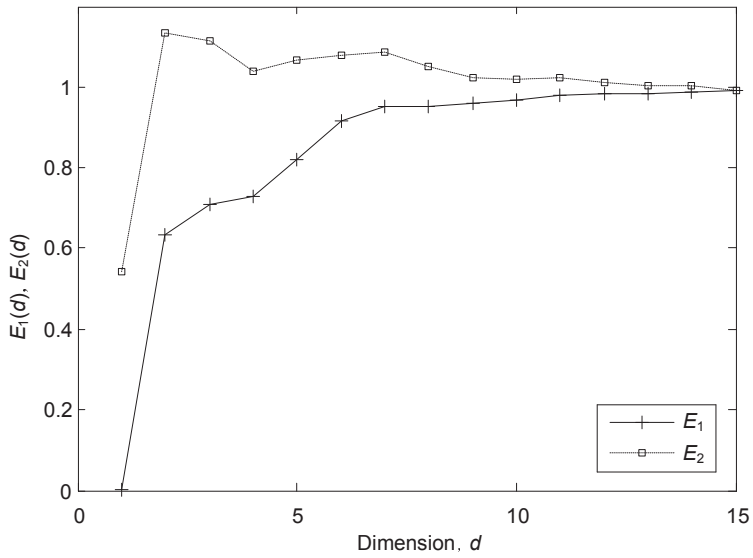

Fig. 15. The values of E_1 and E_2 of envelope acceleration data.

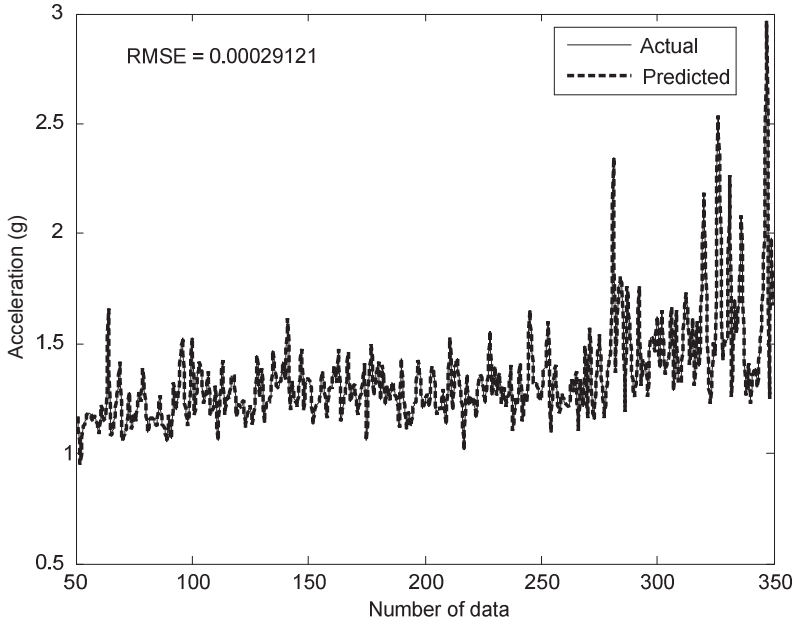

Fig. 16. Training and validating results of envelope acceleration data.

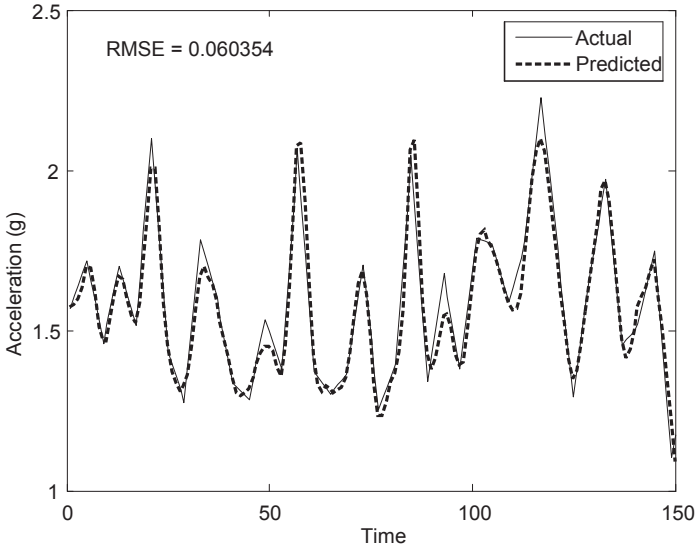

Fig. 17. Predicted results of envelope acceleration data.

Table 1 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