A Classification and Regression Trees (CART) Model of Parallel Structure and Long-term Prediction Prognosis of Machine Condition

Van Tung Tran[1], Bo-Suk Yang[1,*], Andy Chit Chiow Tan[2]

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

[2]*School of Engineering Systems, Queensland University of Technology, G.P.O. Box 2343, Brisbane, Qld. 4001, Australia*

This paper presents a combined prediction model involving the parallel of classification and regression trees (CART) model, namely p-CART, and a long-term direct prediction methodology of time series techniques to predict the future stages of the machine's operating conditions. p-CART model consists of multiple CART models which are connected in parallel. Each sub-model in the p-CART is trained independently. Based on the observations, these sub-models are subsequently used to predict the future values of the machine's operating conditions separately with the same embedding dimension but the different observations' indices. Finally, the predicted results of sub-models are combined to produce the final results of the predicting process. Real trending data acquired from condition monitoring routine of low methane compressor are employed for evaluating the proposed method. A comparative study of the predicted results obtained from traditional CART and p-CART models is also carried out to appraise the prediction capability of proposed model. In addition, a further improvement in predicting capability of p-CART is proposed to ameliorate the accuracy and efficiency of this method.

Keywords: Machine fault prognosis · Long-term time series prediction · CART · Direct prediction methodology.

1 Introduction

Prognosis is known to be a key component in condition-based maintenance (CBM) and has been attracted a great deal of research in recent time due to two following advantages: (i) it increases the safety of machine by detecting the symptoms of failure, so that the system operators can continuously monitor and inspect the stage of failure and take essential and effective maintenance actions on the component before a catastrophic failure occurs; (ii) it reduces the costs of unnecessary maintenance, avoids unplanned breakdown and enables maintenance action to be scheduled more effectively. Nevertheless, prognosis has been still a

difficult task of CBM.

According to [1], prognosis techniques can be broadly classified into three categories: probability-based, model-based, and data-driven based. Among techniques, model-based and data-driven based are the most considered because they provide higher accuracy and reliability. *Model-based prognosis techniques* are applicable in situations where accurate mathematical models can be constructed based on the physical fundamentals of the system; or the models require available extensive failure data to which are either too costly or impossible to be obtained. Some of the published researches using these techniques can be found in references [2-6]. Even though the accuracy of these techniques is reasonably high, they are only suitable for specific components and each component requires a specific mathematical model. Alternatively, *data-driven prognosis techniques* require a large amount of historical failure data to build a prognostic model that learns the system behavior. They frequently use vibration signals for temporal pattern identifications since it is relatively easy to measure and record machine vibration data. Therefore, data-driven prognosis techniques with vibration-based measurement have been developed in the recent time. Some salient researches using these techniques have been proposed in references [7-11].

In condition prognosis, the prediction model uses the available observations to forecast the future operating conditions of the machine [10]. From these predicted results, remaining useful life (RUL) of machine can be prognosticated. RUL is the time interval between the current operating condition point and the point where the predicted values fall within the alarm region or reach the predetermined failure threshold. Consequently, the more the future operating conditions of machine are accurately predicted, the easier RUL is determined. Hence, long-term prediction is essential for machine condition prognosis even though it is still a difficult and challenging task in time series prediction domain [12].

In long-term prediction, embedding dimension (ED), time delay (TM), and selection prediction model are essential to be considered. ED and TM are used to reconstruct the space state of machine's condition time series and establish the fundamental parameters of prediction model. ED is the number of initial observations that should be used as the inputs for prediction model. This value can be determined by using the false nearest neighbor method (FNN) [13] or the Cao's method [14]. TM is the number of steps that can be predicted by the prediction model to obtain the optimum performance. It can be calculated by using some of the published methods such as auto-correlation [15], average displacement [16], and auto-mutual information [17] which is used in this study. The next problem in prognostic system is the selection of prediction model. Classification and regression tree (CART) [18] has been widely implemented in machine fault diagnosis. In the aspect of prediction, CART as well as its extensions has been

applied to forecast short-term load of power system [19, 20] and operating condition of machine [21] with memorable performance. However, these researches merely focused on short-term prediction methodology.

This study proposes a combined prediction model in which multiple CART models, namely p-CART, are connected in parallel for long-term prediction purpose. Each sub-model in the p-CART is trained independently. Based on the observations, these sub-models are then used to predict the future values of the machine's operating conditions separately with the same embedding dimension but the different observations' indices. Finally, the predicted results of sub-models are combined to produce the final results of the predicting process. The parallel-structure model in general and p-CART model in particular has many advantages. It can use several observations simultaneously to enhance the prediction accuracy. This is not suitable for the traditional CART because the more observations lead to an expansion in embedding dimensions that can result in the increase of computational complexity. The parallel-structure model has been applied for forecasting purposes as indicated in references [22-23]. However, these researches merely addressed the short-term prediction method.

2 Background Knowledge

2.1 Time Delay (TM) Estimation

There are several methods published in literature could be used to choose the TM. However, most of them are based on empirical concepts and is not easy to identify which of the methods is suitable for a particular task. In this paper, TM is dealt with auto mutual information (AMI) method. The mutual information (MI) can be used to evaluate the dependence among random variables. The MI between two variables, let X and Y be the amount of information obtained from X in the presence of Y and vice versa. In time series prediction problem, if Y is the output and X is a subset of the input variables, the MI between X and Y is one criterion for measuring the dependence between inputs and output. Thus, the inputs subset X, which gives maximum MI, is chosen to predict the output Y. The MI between two measurements taken from a single time series $x(t)$ separated by time τ is called the AMI. The detailed theory of AMI was presented in references [17, 24-25]. AMI estimates the degree to which the time series $x(t+\tau)$ on average can be predicted from a given time series $x(t)$, i.e. the mean predictability of future values in the time series from the past values.

The AMI between $x(t)$ and $x(t+\tau)$ is:

$$I_{XX_\tau} = \sum_{x(t),x(t+\tau)} P_{XX_\tau}(x(t), x(t+\tau)) \ln\left(\frac{P_{XX_\tau}(x(t), x(t+\tau))}{P_X(x(t)) P_{X_\tau}(x(t+\tau))}\right) \quad (1)$$

where $P_X(x(t))$ is the normalized histogram of the distribution of values observed for $x(t)$ and $P_{XX_\tau}(x(t), x(t+\tau))$ is the joint probability density for the measurements of $x(t)$ and $x(t+\tau)$.

The decreasing rate of the AMI with increasing time delay is a normalized measure of the time series' complexity. The first local minimum of the AMI of time series has been used to determine the optimal TM.

2.2 Determining the Embedding Dimension (ED)

After calculating the TM, ED is the next parameter to be determined. FNN method is employed in this study and will be briefly explained. Assuming that a time-series of $x_1, x_2, ..., x_N$ and vector $y_i(d)$, which is given in equation (2), in a delay coordinate embedding of the time series with time delay τ and embedding dimension d are given.

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

The observations x_i are projections of the system's trajectory in the multivariate state space onto 1-dimensional axis. The FNN method is based on the concept that in the passage from dimension d to dimension d+1, where one can differentiate between points which are 'true" or "false" neighbor on the orbit. For instance in Figure 1, points A, B, C and D belong to a curve. In 1-dimension, points A and D appear to be nearest neighbor. However, point D is no longer nearest neighbor of point A in 2-dimension. In the same way, points A and C are nearest neighbor in 2-dimension but they are no longer neighbors when viewed in 3-dimension. In this case, points A, D, C are examples of "false" neighbors while points A and B are "true" neighbors.

Figure 1 An example of false nearest neighbors

The criteria for identification of false nearest neighbors can be explained as follows: denote $y_i^r(d)$ as the nearest neighbor of $y_i(d)$ in a d dimensional embedding space. According to [13], the nearest neighbor is determined by finding the vector which minimizes the Euclidean distance:

$$R_d = \left\| y_i(d) - y_i^r(d) \right\| \tag{3}$$

Considering each of these vectors under a d+1 dimensional embedding:

$$y_i(d+1) = [x_i, x_{i+\tau}, x_{i+2\tau}, ..., x_{i+d\tau}], \quad i = 1, 2, ..., N-d\tau \tag{4}$$

$$y_i^r(d+1) = [x_i^r, x_{i+\tau}^r, x_{i+2\tau}^r, ..., x_{i+d\tau}^r], \quad i = 1, 2, ..., N - d\tau \tag{5}$$

The vectors are separated by the Euclidean distance:

$$R_{d+1} = \left\| y_i(d+1) - y_i^r(d+1) \right\| \tag{6}$$

The first criterion of FNN which identifies a false nearest neighbor is:

$$\sqrt{\frac{R_{d+1}^2 - R_d^2}{R_d^2}} = \frac{\left| x_{i+d\tau} - x_{i+d\tau}^r \right|}{R_d} > R_{tol} \tag{7}$$

where R_{tol} is a tolerance level.

The second criterion is:

$$\frac{R_{d+1}}{R_A} > A_{tol} \tag{8}$$

where R_A is a measure of the size of the attractor and A_{tol} is a threshold that can be chosen in practice. If both equations (7) and (8) are satisfied, then $y_i^r(d)$ is a false nearest neighbor of $y_i(d)$. Once the total number of FNN is calculated, the percentage of FNN is measured. An appropriate ED is the value where the percentage of FNN falls to zero.

2.3 Regression Trees

CART method has been extensively developed for classification or regression purpose depending on the response variable which is either categorical or numerical. In this study, CART is utilized to build 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.

2.3.1 Tree growing Let L be a learning data which 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 $\mathbf{x}_i$. In order to build the tree, learning data L is recursively partitioned into two subsets by binary split until the terminal nods are achieved. The result is to move the couples $(y, \boldsymbol{x})$ to left or right nodes which contain more homogeneous responses. The predicted response at each terminal node t is estimated by the mean $\bar{y}(t)$ of the n(t) response variables y contained in that terminal node.

The split selection at any internal node t is chosen according to the node impurity that is

measured by within-node sum of squares:

$$R(t) = \frac{1}{n} \sum_{y_i, \boldsymbol{x}_i \in t} (y_i - \bar{y}(t))^2 \tag{9}$$

and,

$$\bar{y}(t) = \frac{1}{n(t)} \sum_{y_i, \boldsymbol{x}_i \in t} y_i \tag{10}$$

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 verifies:

$$\begin{aligned} &\Delta R(s^*, t) = \max \Delta R(s, t), \quad s \in S \\ &\Delta R(s, t) = R(t) - R(t_L) - R(t_R) \end{aligned} \tag{11}$$

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

2.3.2 Tree pruning The tree gained in tree growing process has many terminal nodes that increase the precision of the responses. However, this is frequently too complicated and over-fitting is highly probable. Consequently, it should be pruned back.

Tree pruning process is performed by the following procedure:

Step 1: At every internal node, an error-complexity is found for the number of descendant sub-trees. 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, $\alpha \geq 0$ is the complexity parameter which weights the number of terminal nodes.

Step 2: Using the error-complexity attained in 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.

2.3.3 Cross-validation for selecting the best tree There are two possible methods to select the best tree. One is through the use of independent test data and the other is cross-validation which is used in this study.

The learning data L is randomly divided into v approximately equal group, and (v−1) groups are then utilized as the learning data for growing the tree model. The remaining group is employed as testing data for error estimation of tree model. As a result, v errors are obtained by

v iterations with variation of the combinations of the learning data and testing data. The mean and standard deviation of the errors are given:

$$R^{CV}(d)=\frac{1}{v}\sum_{i=1}^{v}R^{ts}(d_i),\quad \sigma(R^{CV}(d))=\sqrt{\frac{1}{v}\sum_{i=1}^{v}\left\{R^{ts}(d_i)-R^{CV}(d)\right\}^2} \tag{13}$$

where $R^{CV}(\cdot)$ is the average relative error, d is the cross-validation tree, σ is the standard error, and $R^{ts}(\cdot)$ is the testing data error.

The best tree T_t selection is adopted:

$$R(T_t)=R^{CV}(T_{\min})+\sigma(R^{CV}(T_{\min})) \tag{14}$$

where $R(\cdot)$ is the cross-validation error and T_{min} is the tree with the smallest cross-validation error.

3 Long-term Direct Prediction Strategy for p-CART Model

Prediction techniques consist of short-term prediction (one-step ahead prediction) and long-term prediction (multi-step ahead prediction). Unlike the short-term prediction, the long-term prediction is typically faced with growing uncertainties arising from various sources such as the accumulation of errors and the lack of information. According to Sorjamaa et al. [26], long-term prediction is divided into three frequently used strategies that involve recursive prediction, direct prediction and DirRec prediction. The detailed information of these strategies could be found in reference [27]. In this section, the direct prediction strategy applying for p-CART model is specifically presented.

Assuming that a sequence of observation $y_t=[x_{t-d+1},x_{t-d+2},...,x_t]$ is given, to predict h future values $\hat{y}_{t+h}=[x_{t+1},x_{t+2},...,x_{t+h}]$, the H different parallel-structure prediction models are used. These models are generated by using training set D. The training set D including input vectors $\boldsymbol{X}_i$ and output vectors $\boldsymbol{Y}_i$ is created from the given observations $y_t=[x_1,x_2,...,x_t]$ by using a sliding window of length $d\times N+h$, where N is the number of sub-models in parallel-structure prediction model. The vector $\boldsymbol{X}_i$ corresponds to the first $d\times N$ value of window whilst the vector $\boldsymbol{Y}_i$ is the remaining h values of window. The number of elements in each vector $\boldsymbol{X}_i$ is $d\times N$ which is used to generate N sub-models of parallel-structure models. The training set is structured by synthesizing $\boldsymbol{X}$ and $\boldsymbol{Y}$ vectors in the form as shown in Table 1. Thus, by using training set, H parallel-structure prediction models are sequentially generated with different output $\boldsymbol{Y}_i$ which include all the values in the ith column $\boldsymbol{Y}$ in training set D.

Table 1 Training set *D* for direct prediction strategy.

4 Architecture of p-CART model

The p-CART model consists of several sub-models of CART in parallel. Each of these sub-models is independently trained with the same output and input vectors. However, not all the elements of input vectors are used for training because the total number of elements in input vectors is $d \times N$ which is larger than the value of embedding dimension. Therefore, the indices of elements are modified corresponding to each sub-model and total elements used as inputs for each sub-model is equivalent to embedding dimension. For example, the p-CART is used for forecasting the future values in which the number of sub-model N is 3 while ED and TM are calculated as 3 and 4, respectively. Thus, the number of elements of input vectors is 9. The sub-model CART 1 takes the elements x_{t-2}, x_{t-1}, x_t as input. Similarly, the sub-model CART 2 uses the elements $x_{t-5}, x_{t-3}, x_{t-1}$ and the sub-model CART 3 acquires the elements $x_{t-8}, x_{t-5}, x_{t-2}$ as their inputs. The output vector of each sub-model is the same vector as $[x_{t+1}, x_{t+3}, x_{t+3}, x_{t+4}]$. The architecture and input elements for sub-model is shown in Figure 2.

Finally, the predicted values from each sub-model are combined to determine the final predicted results by using average formula:

$$\hat{y}_{t+h} = \left[\frac{1}{N} \sum_{1}^{N} \hat{x}_{t+1}, \frac{1}{N} \sum_{1}^{N} \hat{x}_{t+2}, ..., \frac{1}{N} \sum_{1}^{N} \hat{x}_{t+h} \right] \tag{15}$$

Figure 2 Architecture and input values for sub-model of p-CART

5 Proposed System

The proposed system for prognosis of machine condition comprises four procedures sequentially as shown in Figure 3, namely, data acquisition, data splitting, training-validating model, and predicting. The role of each procedure is explained as follows:

Step 1 Data acquisition: this procedure is used to obtain the vibration data from machine condition. It covers a range of data from normal operation to obvious faults of the machine.

Step 2 Data splitting: the trending data attained from previous procedure is split into two parts: training set and testing set. Different data is used for different purposes in the prognosis system. Training set is used for creating the prediction models whilst testing set is utilized to test the trained models.

Step 3 Training-validating: this procedure includes the following sub-procedures: estimating the TM and determining the ED based on AMI and FNN method, respectively; creating the

prediction models and validating those models. Validating the prediction models are used for measuring their performance capability.

Step 4 Predicting: long-term direct prediction method is used to forecast the future values of machine condition. The predicted results are measured by the error between predicted values and actual values in the testing set. Updating models are also carried out in this procedure for the next prediction process.

Figure 3 Proposed system for machine fault prognosis

6 Experiments and Results

The proposed method is applied to a real system to predict the trending data of a low methane compressor of a petrochemical plant. The compressor shown in Figure 4 is driven by a 440 kW motor, 6600 volt, 2 poles and operating at a speed of 3565 rpm. Other information of the system is summarized in Table 2.

Figure 4 Low methane compressor
Table 2 Information of the system

The condition monitoring system of this compressor consists of two types: off-line and on-line. In the off-line system, accelerometers were installed along axial, vertical, and horizontal directions at various locations of drive-end motor, non drive-end motor, male rotor compressor and suction part of compressor. In the on-line system, accelerometers were located at the same positions as in the off-line system but only in the horizontal direction.

The trending data was recorded from August 2005 to November 2005 which included peak acceleration and envelope acceleration data. The average recording duration was 6 hours during the data acquisition process. Each data record consisted of approximately 1200 data points as shown in Figures 5 and 6, and contained information of machine history with respect to time sequence (vibration amplitude). Consequently, it can be classified as time-series data.

Figure 5 The entire peak acceleration data of compressor
Figure 6 The entire envelope acceleration data of compressor

These figures show that the machine was in normal condition during the first 300 points of the time sequence. After that time, the condition of the machine suddenly changed. This indicates possible faults were occurring in the machine. By disassembling and inspecting, these faults were identified as the damage of main bearings of the compressor (notation Thrust: 7321 BDB) due to insufficient lubrication. Consequently, the surfaces of these bearings were overheated and delaminated [21].

With the aim of forecasting the change of machine condition, the first 300 points were used

to train the system. Before being used to generate the prediction models, the TM was initially calculated according to the method mentioned in section 2.1. Theoretically, the optimal TM is the value at which the first local minimum of the AMI is obtained. From Figure 7, the optimal TM of peak acceleration training data was found as 7. Similarly, 5 was the optimal TM value of envelope acceleration training data.

Figure 7 Time delay estimation.

Using FNN method described in section 2.2, the optimal TM is subsequently utilized to determine the embedding dimension d. It is noted that the tolerance level R_{tol} and threshold A_{tol} must be initially chosen. In this study, R_{tol} = 15 and A_{tol} = 2 were used according to [13]. The relationship between the false nearest neighbor percentage and the embedding dimension for both peak acceleration data and envelope acceleration data was shown in Figure 8. From the figure, the embedding dimension d was chosen as 4 for both dataset where the false nearest neighbor percentage reaches 0.

Figure 8 The relationship between FNN percentage and embedding dimension.

Subsequent to determining the TM and ED, the process of generating the prediction models was carried out. Based on those values, the training data was created in which the number of elements of input vectors is equal to the product of the ED and the number of sub-models; and the number of elements of output vectors is equal to the TM. The process of creating the training data has been described in section 3. Based on the training data, the p-CART model was established. It is noted that during the process of building each sub-model, the number of response values for each terminal node in tree growing process was 5 and 10 cross-validations were decided for selecting the best tree in tree pruning. Furthermore, in order to evaluate the predicting performance, the root-mean square error (RMSE) was utilized as follows,

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2}{N}} \tag{16}$$

where N, y_i, $\hat{y}_i$ represent the total number of data points, the actual value and predicted value of prediction model in the training data or testing data, respectively.

Choosing the number of sub-models for the p-CART model is another problem to be considered in order to obtain the best predicting performance. However, it is not easy to identify how many sub-models should be used and has not been extensively reported in literature. Therefore, practical approach is a viable solution to deal with this problem. In this paper, the number of sub-models was examined from 2 to 5. Then, the predicted results were compared to

determine how many sub-models are appropriate for obtaining the best performance. Table 3 showed a summary of the RMSE values of peak acceleration and envelope acceleration data. From this table, all the RMSEs of the p-CART were smaller than those of traditional CART in both cases of peak acceleration and envelope acceleration, especially in the training process. The more the sub-models were, the smaller RMSE value was obtained from the training result. However, this was inappropriate to the testing process. The smallest RMSE value was attained when the number of sub-models was 3. After that, the RMSE gradually increased. The reason could be related to the expansion in observations due to the increase in the number of sub-models. This expansion leads to the sub-model using improper observations in prediction process.

Table 3 The RMSEs of CART and p-CART

Due to the best performance of p-CART obtained from 3 sub-models, a comparison of tracking capability of the operating condition change between this model and traditional model was carried out. The training results of the CART models for peak acceleration and envelope acceleration data were respectively performed and shown in Figures 9(a) and 10(a). The actual and predicted values were almost identical with very small RMSE values of 0.002217 and 1.314×10^{-5} for acceleration data and envelope data, respectively. Similarly, the training results of p-CART were depicted in Figures 9(b) and 10(b). The values of RMSE were 0.0007 for acceleration data and 7.088×10^{-6} for envelope data, which were significantly smaller than those of CART. These results indicate that the proposed method manifests the advances in learning capability.

Figure 9 Training and validating results of peak acceleration data
Figure 10 Training and validating results of envelope acceleration data

Figures 11 and 12 showed the predicted results in testing process of the CART and p-CART models for peak acceleration and envelope acceleration. The RMSE values obtained from p-CART for both data were smaller than those of CART. Furthermore, in Figure 11, the p-CART was superior to CART model in keeping track of the changes of the machine operating condition. This indicated that there was an improvement in the predicting ability of p-CART model when compared with that of traditional model.

Figure 11 Predicted results of peak acceleration data.
Figure 12 Predicted results of envelope acceleration data.

5 Conclusions

This paper presents a data-driven approach which combines the prediction model involving

parallel of CART and long-term direct prediction methodology for forecasting the operating conditions of machine. The p-CART model is validated for its ability to predict future state conditions of a low methane compressor using peak acceleration and envelope acceleration data. The predicted results of the p-CART are also compared with those of traditional CART. From the predicted results, the p-CART model performance is superior to the traditional model. Consequently, this proposed method is a new way to improve the performance of long-term prediction which has been a difficult and a challenging task in forecasting the machine's operating conditions.

Furthermore, this study commences the new feasibility to further improve the predicting capability of p-CART. This improvement is implemented by multiplying the predicted result of each sub-model by the weight factor when calculating the final predicted results. These weight factors are determined based on the errors during training process. The sub-model which has the most approximate predicted values to the actual values will be assigned with the largest weight factor while the others will be assigned with descendent weight factors. The final predicted result will be the sum of these products. This improvement proposed here is mentioned as a future work.

References

1. Vachtsevanos, G., Lewis, F., Roemer, M., Hess, A. and Wu, B. (2006). *Intelligent fault diagnosis and prognosis for engineering system*. Wiley, New York.
2. Tu, F., Ghoshal, S., Luo, J., Biswas, G., Mahadevan, S., Jaw, L. and Navarra, K. (2007). PHM integration with maintenance and inventory management systems. *Proceedings of Aerospace Conference 2007 IEEE*, pp. 1–12.
3. Abbas, M., Ferri, A.A., Orchard, M.E. and Vachtsevanos, G.J. (2007). An intelligent diagnostic/ prognostic framework for automotive electrical systems. *Proceedings of Intelligent Vehicles Symposium 2007 IEEE*, pp. 352–357.
4. Watson, M., Byington, C., Edwards, D. and Amin, S. (2005). Dynamic modeling and wear-based remaining useful life prediction of high power clutch systems. *Tribology Transactions*, *48*, 208–217.
5. Luo. M., Wang, D., Pham, M., Low, C.B., Zhang, J.B., Zhang, D.H. and Zhao, Y.Z. (2005). Model-based fault diagnosis/prognosis for wheeled mobile robots: a review. *Proceeding of 32nd Annual Conference of IEEE*, *Industrial Electronics Society*, pp. 2267–2272.
6. Li, Y., Kurfess, T.R. and Liang, S.Y. (2000). Stochastic prognostics for rolling element bearings. *Mechanical Systems and Signal Processing*, *14*, 747–762.
7. Schwabacher, M. and Goebel, K. (2007). An survey of artificial intelligence for prognostics. *Proceedings of AAAI Fall Symposium on Artificial Intelligence for Prognostics*, 9-11 Nov.
8. Vachtsevanos, G. and Wang, P. (2001). Fault prognosis using dynamic wavelet neural networks. *AUTOTESTCON Proceedings, IEEE Systems Readiness Technology Conference*, pp. 857–870.
9. Huang, R., Xi, L., Li, X., Liu, C.R., Qiu, H. and Lee, J. (2007). Residual life prediction for ball bearings based on self-organizing map and back propagation neural network methods. *Mechanical Systems and Signal Processing*, *21*, 193–207.
10. Wang, W.Q., Golnaraghi, M.F. and Ismail, F. (2004). Prognosis of machine health condition using neuro-fuzzy system. *Mechanical System and Signal Processing*, *18*, 813–831.
11. Brown, E.R., McCollom, N.N., Moore, E. and Hess, A. (2007). Prognostics and health management – a data-driven approach to supporting the F-35 Lightning II. *Proceedings of Aerospace Conference 2007 IEEE*, pp. 1–12.

12. Ji, Y., Hao, J., Reyhani, N. and Lendasse, A. (2005). Direct and recursive prediction of time series using mutual information selection. *Lecture Notes in Computer Science*, *3512*, 1010–1017.
13. Kennel, M.B., Brown, R. and Abarbanel, H.D.I. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, *45*, 3403–3411.
14. Cao, L. (1997). Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D*, *110*, 43–50.
15. Broomhead, D.S. (1986). Extracting qualitative dynamics from experimental data. *Physica D*, *20*, 217.
16. Rosenstein, M.T., Collins, J.J. and Luca, C.J.D. (1994). Reconstruction expansion as a geometry-based framework for choosing proper delay time. *Physica D*, *73*, 82–89.
17. Fraser, A.M. and Swinney, H.L. (1986). Independent coordinates for strange attractors from mutual information. *Phys. Rev. A*, *33*, 1134.
18. Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984). *Classification and regression trees*. Chapman & Hall.
19. Yang, J. and Stenzel, J. (2006). Short-term load forecasting with increment regression tree. *Electric Power Systems Research*, *76*, 880–888.
20. Moti, H., Kosemura, N., Ishiguro, K. and Kondo, T. (2001). Short-term load forecasting with fuzzy regression tree in power systems. *System Man and Cybernetics, IEEE International Conference*, 3, pp. 1948–1953.
21. Tran, V.T., Yang, B.S., Oh, M.S. and Tan, A.C.C. (2008). Machine condition prognosis based on regression trees and one-step-ahead prediction. *Mechanical Systems and Signal Processing*, *22*, 1179–1193.
22. Hu, C. and Cao, L. (2007). ANN based load forecasting: a parallel structure. *Systems, Man and Cybernetics, IEEE International Conference*, pp. 3594–3598.
23. Kim, M.S. and Chung, C.S. (2005). Sunspot time series prediction using parallel-structure fuzzy system. *Lecture Notes in Computer Science*, *3614*, 731–741.
24. Jeong, J., Gore, J.C. and Peterson, B.S. (2001). Mutual information analysis of the EEG in patients with Alzheimer's disease. *Clinical Neurophysiology*, *112*, 827–835.
25. Sorjamaa, A., Hao, J., Reyhani, N., Ji, Y. and Lendasse, A. (2007). Methodology for long-term prediction of time series. *Neurocomputing*, *70*, 2861–2869.
26. Sorjamaa, A. and Lendasse, A. (2007). Time series prediction as a problem of missing values: application to ESTSP and NN3 competition benchmarks. Proceedings of European Symposium on Time Series Prediction, pp. 165–174.
27. Tran, V.T., Yang, B.S., and Tan, A.C.C. (2009). Multi-step ahead direct prediction for the machine condition prognosis using regression trees and neuro-fuzzy systems. *Expert Systems with Applications*, 36, 9378-9387.

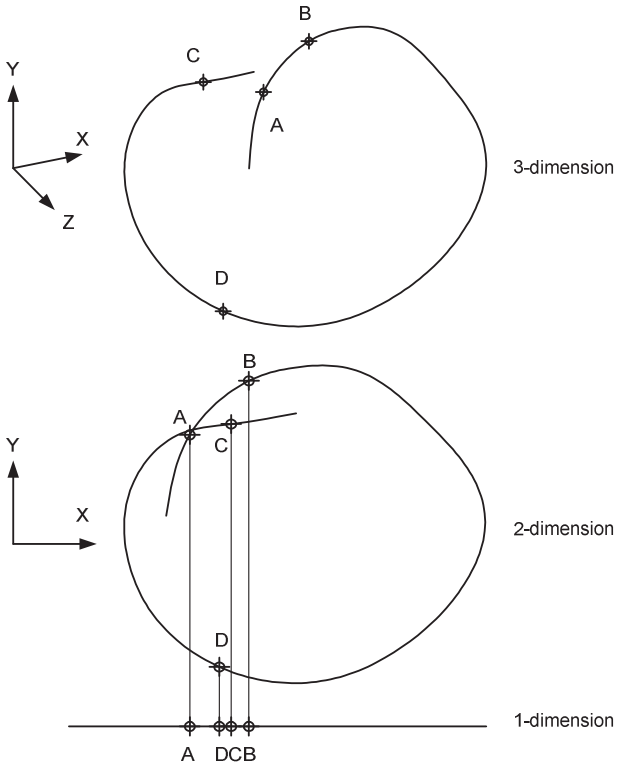


Figure 1 An example of false nearest neighbors

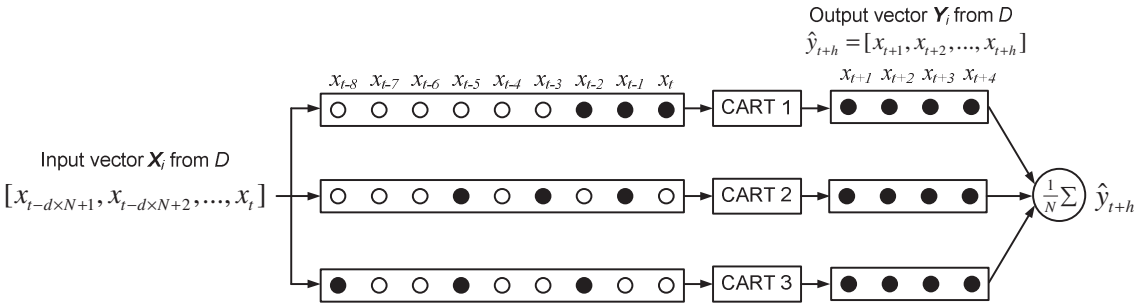


Figure 2 Architecture and input values for sub-model of p-CART.

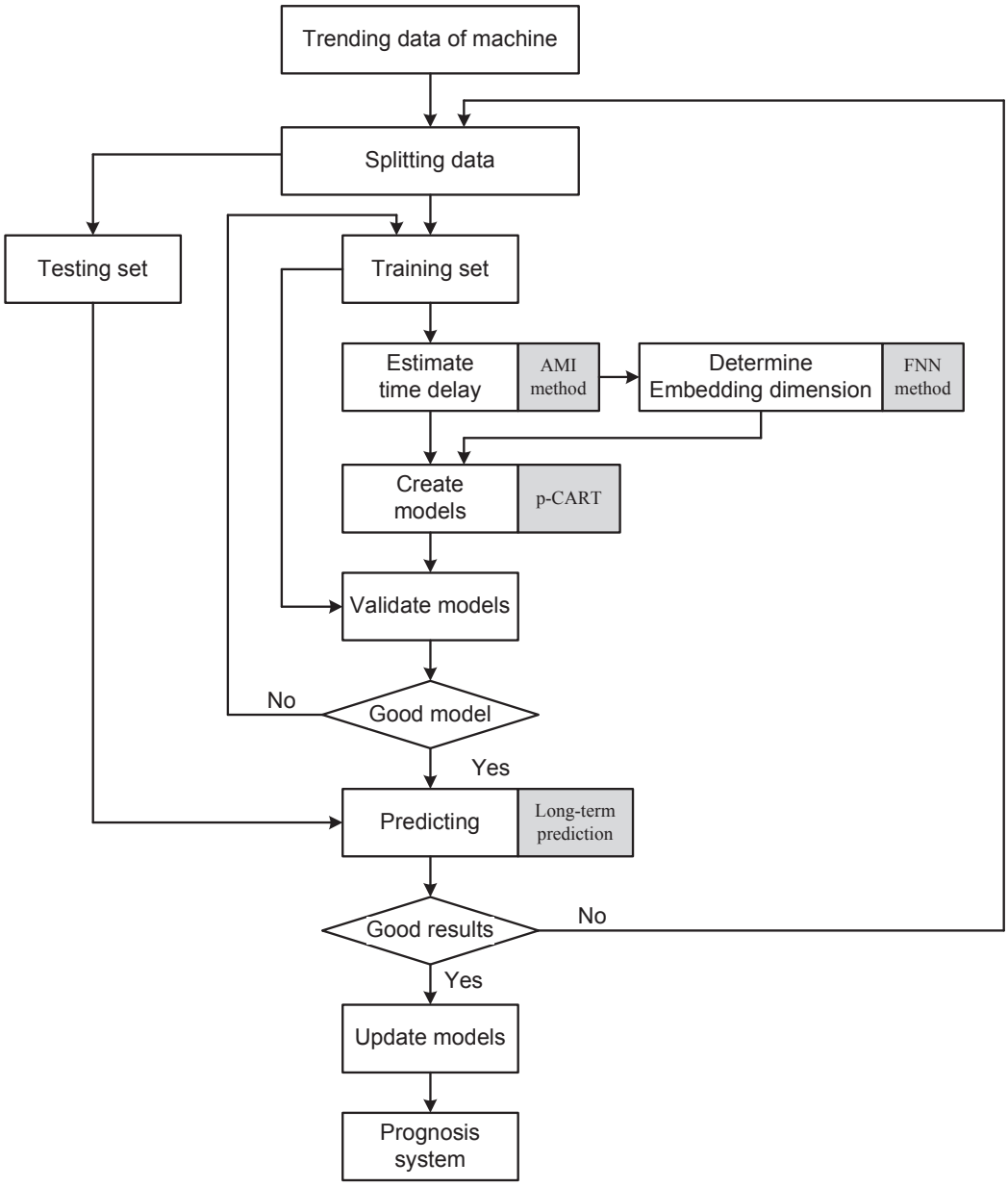


Figure 3 Proposed system for machine fault prognosis.

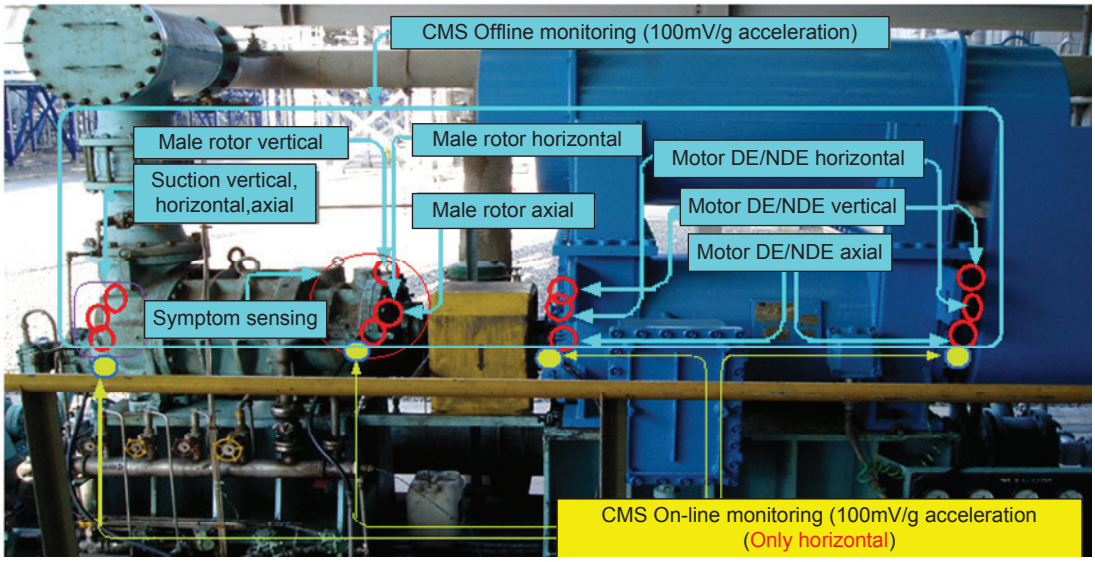


Figure 4 Low methane compressor.

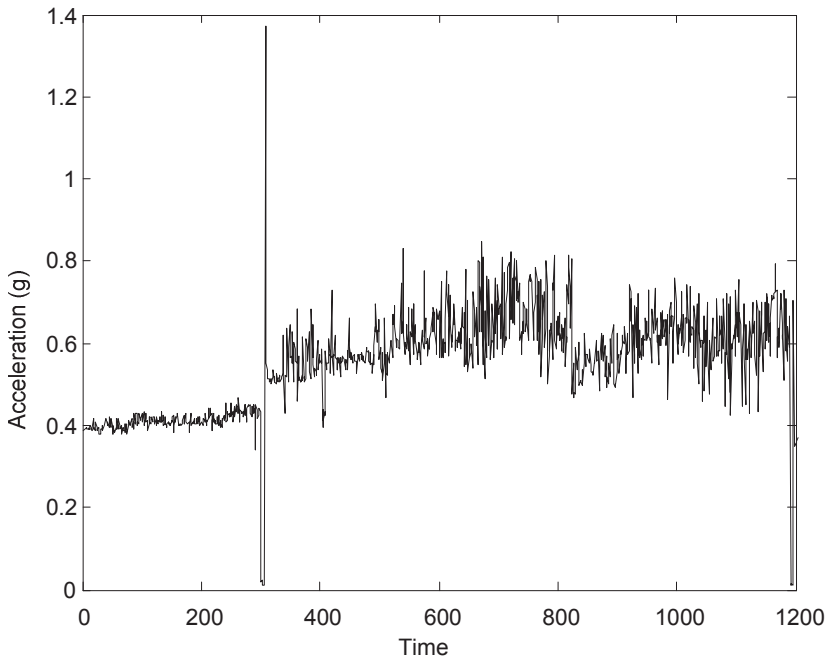


Figure 5 The entire peak acceleration data of compressor.

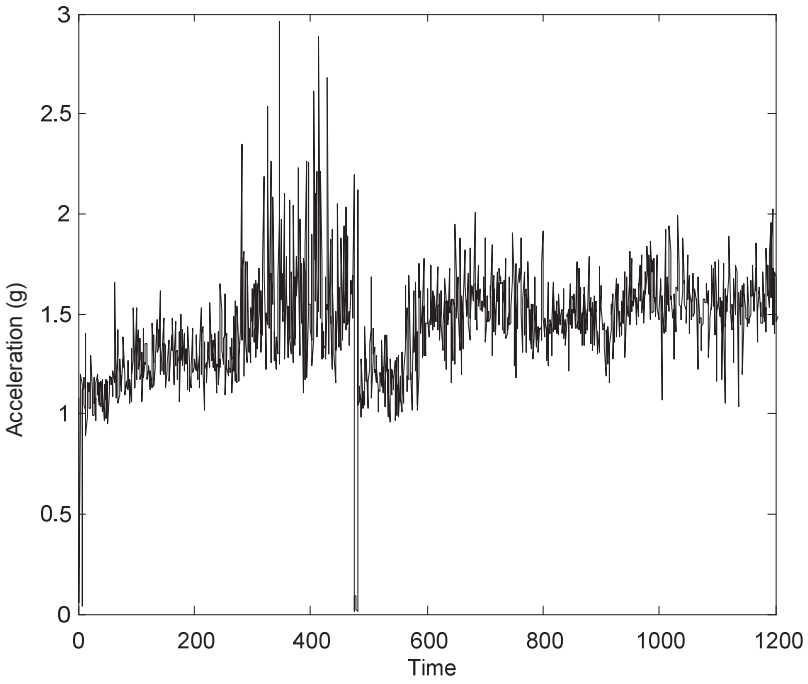


Figure 6 The entire envelope acceleration data of compressor.

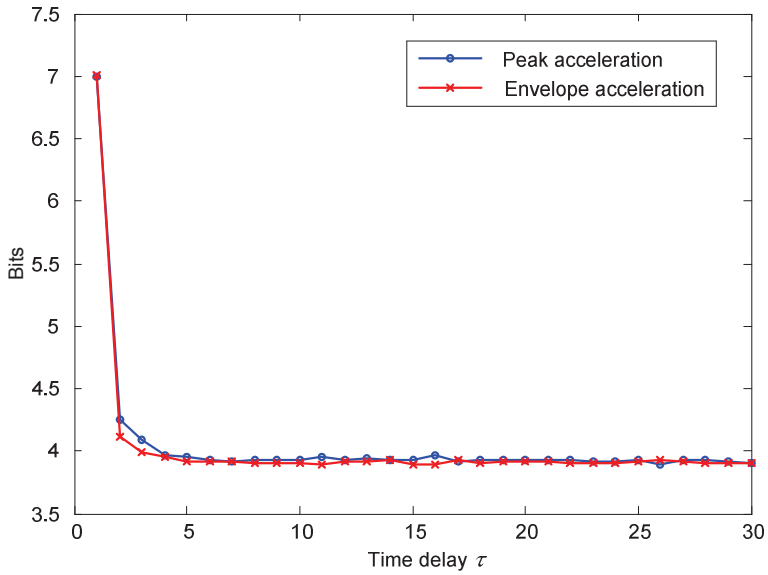


Figure 7 Time delay estimation.

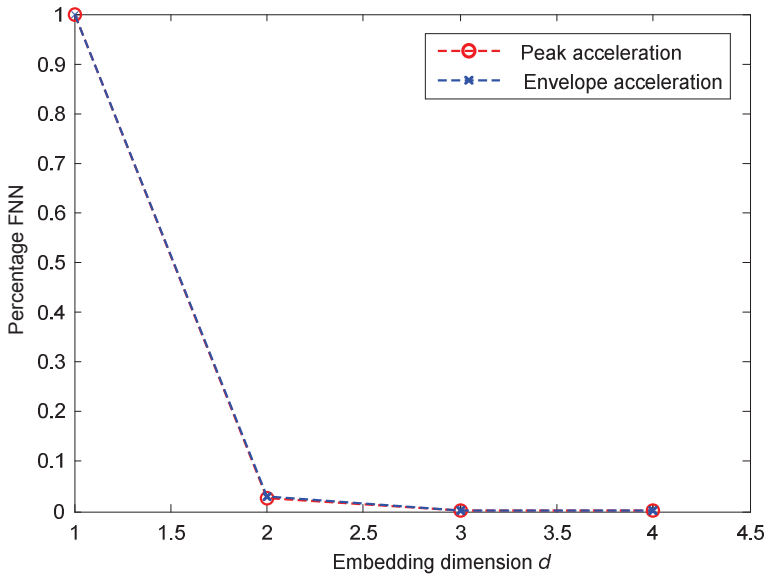
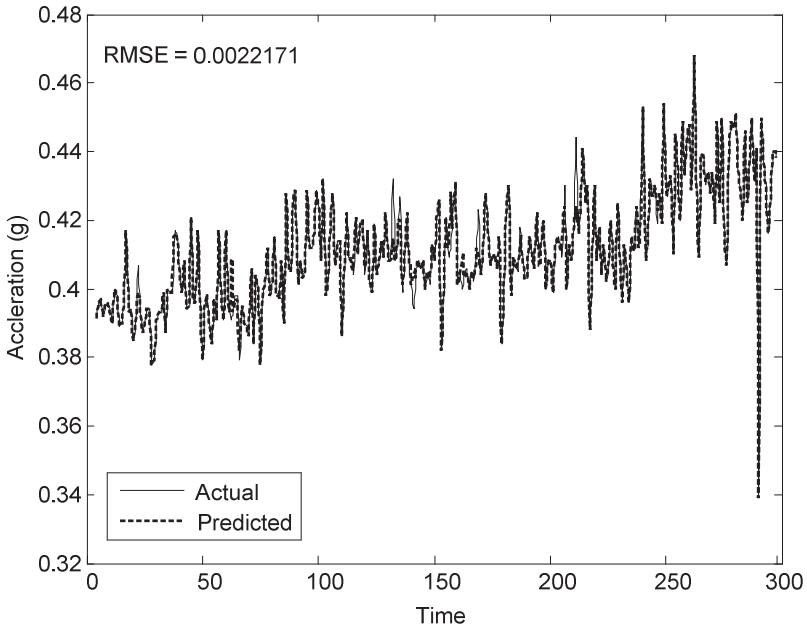
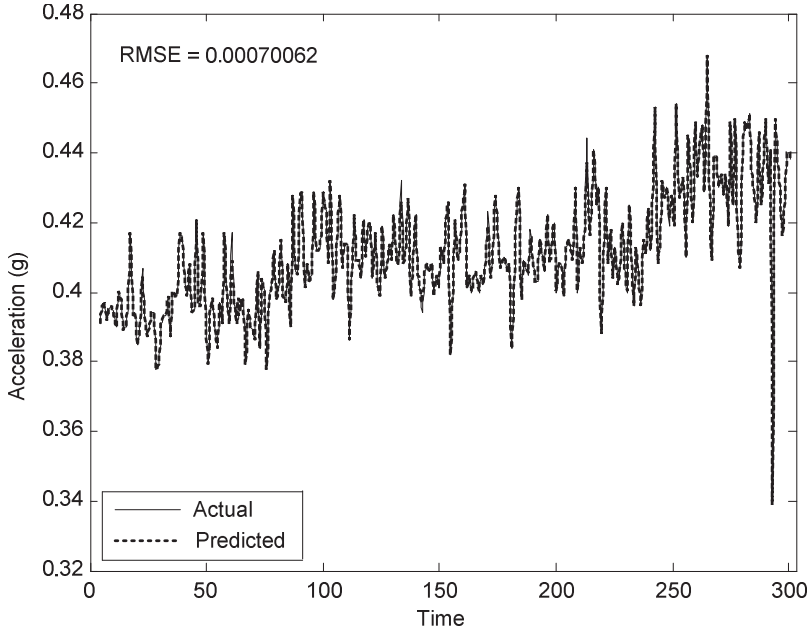


Figure 8 The relationship between FNN percentage and embedding dimension.

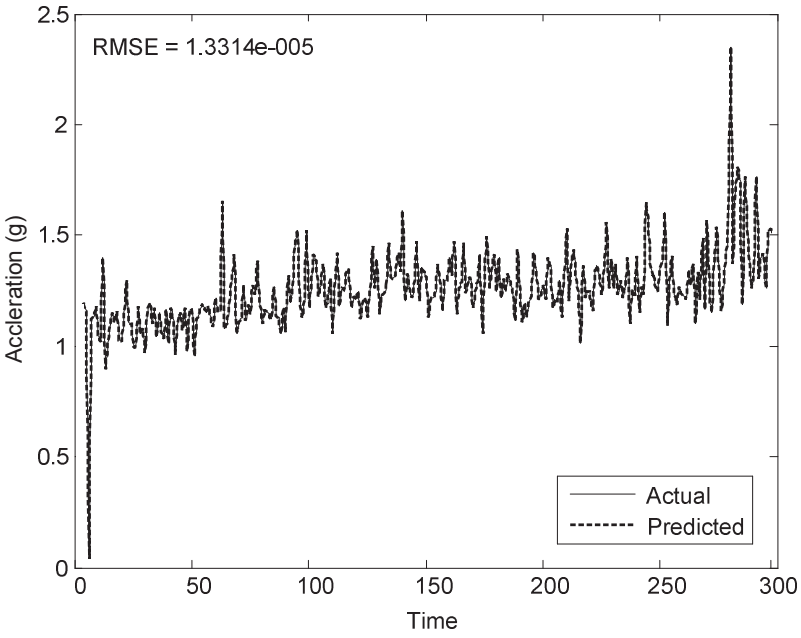


(a)

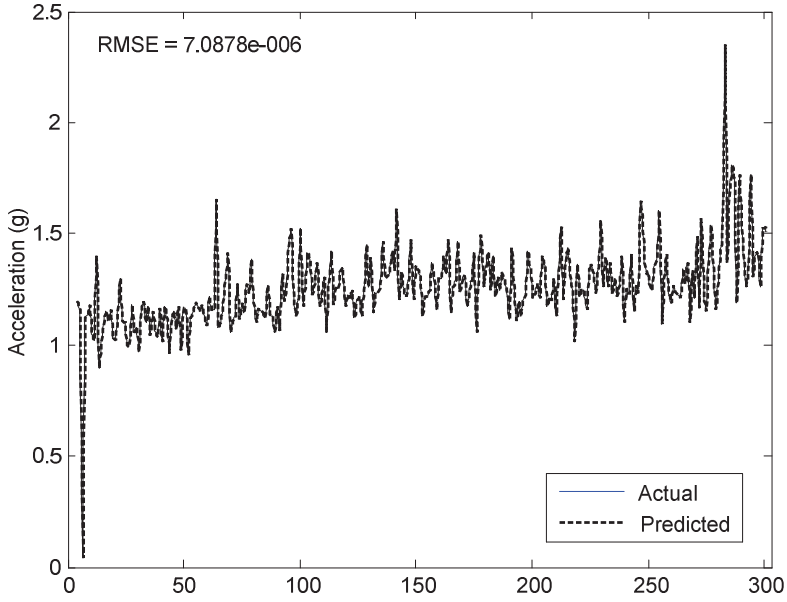


(b)

Figure 9 Training and validating results of peak acceleration data. (a) CART, (b) p-CART.

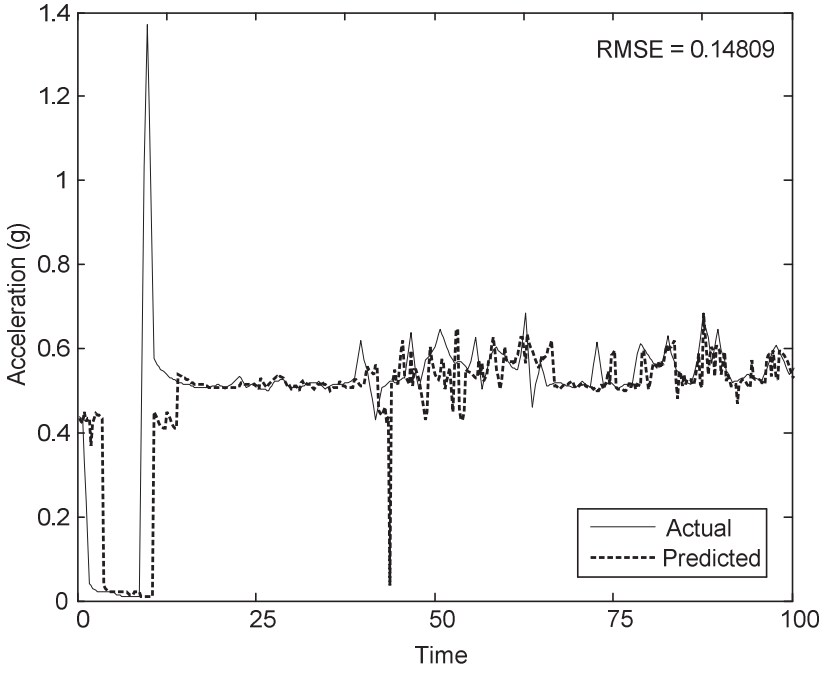


(a)

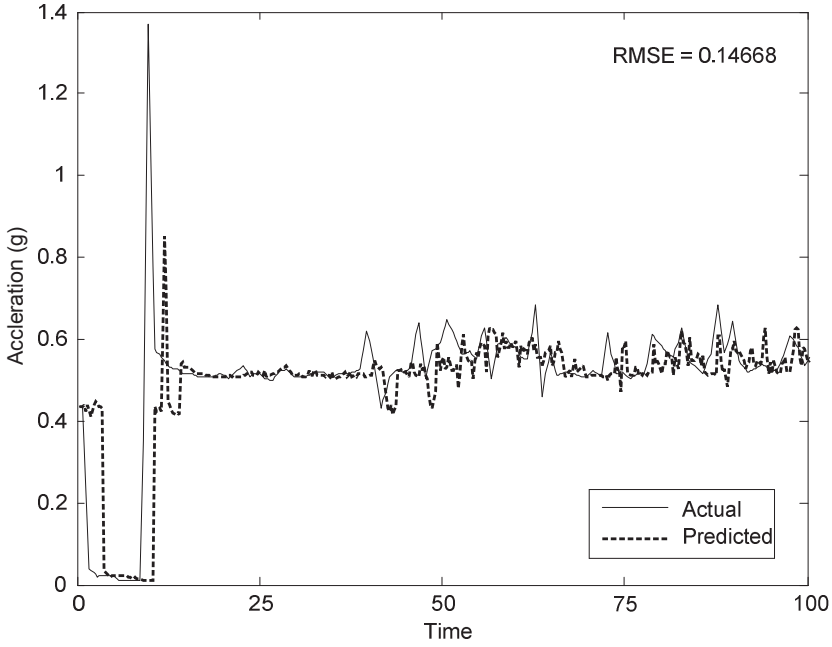


(b)

Figure 10 Training and validating results of envelop acceleration data. (a) CART, (b) p-CART.

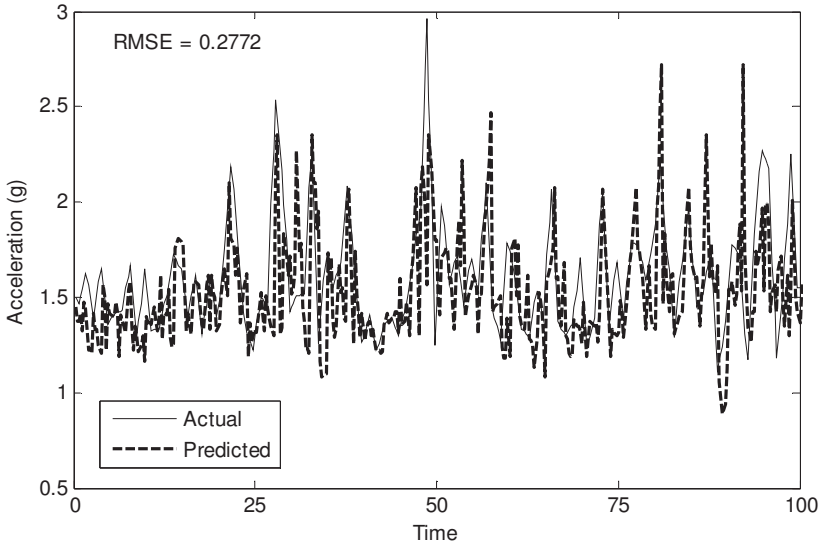


(a)

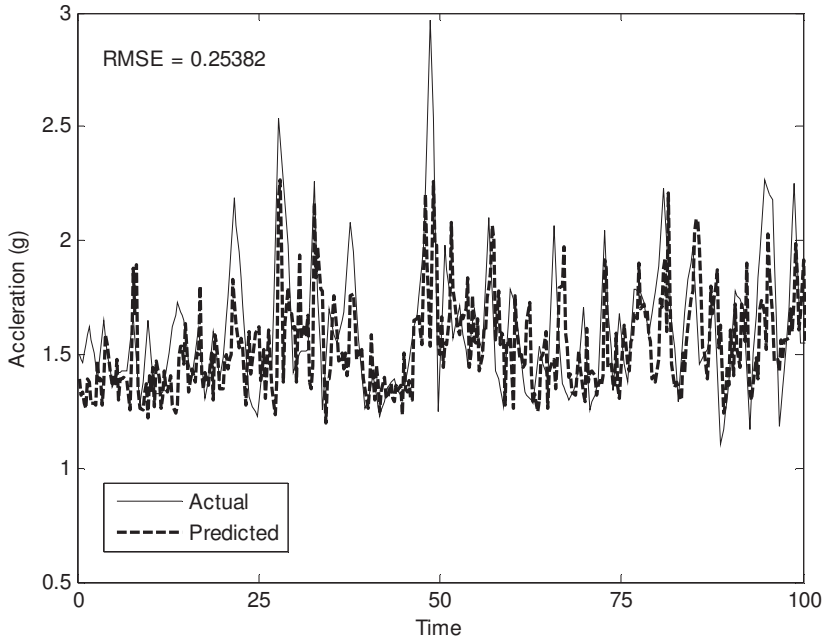


(b)

Figure 11 Predicted results of peak acceleration data. (a) CART, (b) p-CART.



(a)



(b)

Figure 12 Predicted results of envelop acceleration data. (a) CART, (b) p-CART.

Table 1 Training set *D* for direct prediction strategy.

Input $\boldsymbol{X}=[\boldsymbol{X}_1,\boldsymbol{X}_2,...,\boldsymbol{X}_{d\times N}]$	*Output* $\boldsymbol{Y}=[\boldsymbol{Y}_1,\boldsymbol{Y}_2,...,\boldsymbol{Y}_h]$
$[x_1, x_2, ..., x_{d\times N}]$	$[x_{d\times N+1}, x_{d\times N+2}, ..., x_{d\times N+h}]$
$[x_2, x_3, ..., x_{d\times N+1}]$	$[x_{d\times N+2}, x_{d\times N+3}, ..., x_{d\times N+h+1}]$
…	…
$[x_{t-h-d\times N+1}, x_{t-h-d\times N+2}, ..., x_{t-h}]$	$[x_{t-h+1}, x_{t-h+2}, ..., x_t]$

Table 2 Information of the 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

Table 3 The RMSEs of CART and p-CART.

Data	*No. of sub-model*	*CART*		*p-CART*	
		Training	*Testing*	*Training*	*Testing*
Peak acceleration	1	0.002217	0.14809		
	2			0.000923	0.14699
	3			0.000700	0.14668
	4			0.000531	0.14768
	5			0.000425	0.14807
Envelope acceleration	1	1.3314×10^{-5}	0.27720		
	2			8.534×10^{-6}	0.25640
	3			7.088×10^{-6}	0.25382
	4			6.837×10^{-6}	0.26252
	5			5.42×10^{-6}	0.26886