Control and Cybernetics

vol. **39** (2010) No. 1

Machine fault diagnosis and condition prognosis using classification and regression trees and neuro-fuzzy inference systems*

by

Van Tung Tran[1] **and Bo-Suk Yang**[2]

[1] Machine Design Department, Hochiminh City University of Technology Hochiminh, Vietnam

[2] School of Mechanical Engineering, Pukyong National University San 100, Yongdang-dong, Nam-gu, Busan 608-739, South Korea e-mail: bsyang@pknu.ac.kr

Abstract: This paper presents an approach to machine fault diagnosis and condition prognosis based on classification and regression trees (CART) and neuro-fuzzy inference systems (ANFIS). In case of diagnosis, CART is used as a feature selection tool to select pertinent features from data set, while ANFIS is used as a classifier. The crisp rules obtained from CART are then converted to fuzzy if-then rules, employed to identify the structure of ANFIS classifier. The hybrid of back-propagation and least squares algorithm are utilized to tune the parameters of the membership functions. The data sets obtained from vibration signals and current signals of the induction motors are used to evaluate the proposed algorithm. In case of prognosis, both of these models in association with direct prediction strategy for long-term prediction of time series techniques are utilized to forecast the future values of machine operating condition. In this case, the number of available observations and the number of predicted steps are initially determined by false nearest neighbor method and auto mutual information technique, respectively. These values are subsequently utilized as inputs for prediction models. The performance of the proposed prognosis system is then evaluated by using real trending data of a low methane compressor. A comparative study of the predicted results obtained from CART and ANFIS models is also carried out to appraise the prediction capability of these models. The results of the proposed methods in both cases indicate that CART and ANFIS offer a potential for machine fault diagnosis and for condition prognosis.

Keywords: fault diagnosis, classification, induction motors, decision trees, forecasts, fuzzy systems.

*Submitted: January 2009; Accepted: May 2009

1. Introduction

The fault progression process of mechanical systems usually consists of a series of degraded states due to component wear and fatigue during the operating process. Early detection of incipient faults and foretelling the future states of mechanical systems can minimize the costs of unnecessary maintenance, avoid unplanned breakdown and enable maintenance actions to be scheduled more effectively. Thence, the availability and reliability of machine can be increased. Consequently, machine fault diagnosis and machine condition prognosis have been important subjects of research in the recent years.

1.1. Machine diagnosis of induction motors

Machine fault diagnosis is the ability to detect fault, isolate failed component, and decide on potential impacts of the failed component on the system health. Due to the costs of implementing, only critical machine components, whose failures drastically affect the breakdown are frequently examined. In this study, induction motors are considered, due to their indispensable role in several industrial applications. The faults of induction motors may not only cause interruption of product operation, but also increase costs, decrease product quality and effect safety of operators. Consequently, fault diagnosis in induction motors has been the subject of serious studies in the recent years.

The most common faults of induction motors are listed as bearing failures, stator phase winding failures, broken rotor bar or cracked rotor endrings and air-gap irregularities (Acosta,Verucchi and Gelso, 2006). In order to detect/diagnose these faults, system identification and parameter estimation (Sood, Fahs and Henein, 1985; Isermann, 1984; Isermann and Freyermuth, 1991; Cho, Lang and Umans, 1992), as well as other techniques (Yang and Kim, 2006; Yang et al., 2004; Casimir et al., 2006; Widodo, Yang and Han, 2007) have been proposed. These techniques required expensive equipment or accurate mathematical models, which are challenging, to describe the faults of motors. On the contrary, fuzzy logic and artificial neural networks (ANNs) can be used to provide inexpensive but effective fault detection mechanism alternatives (Goode and Chow, 1995). ANNs are good at mapping non-linear numerical information between inputs and outputs. However, ANNs are not interpretable and understandable, i.e. they are incapable of explaining a particular decision to the user in a human-comprehensible form. Conversely, fuzzy logic has the ability of modeling human knowledge in the form of if-then rules by using easily understandable linguistic term. In the case of machine fault diagnosis, involving high levels of uncertainty, due to the complexity of machine systems and unexpected disturbance and noise in sensing (Chen, 1995), fuzzy logic can handle situations, where the answer lies somewhere in-between. Some fuzzy logic based diagnosis approaches for induction motors have been proposed in Satish and Samar (2005), Benbouzid and Nejjari (2001), or Shukri et al. (2004). Nevertheless, the if-then rules, as well as the initial parameters of membership functions, are normally prepared by an expert. Thus, fuzzy logic requires a fine-tuning in order to obtain acceptable rule base and optimal parameters for available data (Shukri et al., 2004). The individual problems from fuzzy logic or ANN, when used alone, can be solved by the integration of both methods, as this has been applied for motor fault diagnosis (Goode and Chow, 1995).

The adaptive neuro-fuzzy inference system (ANFIS), Jang (1993), is an integration of the ANNs adaptive capability and the fuzzy logic qualitative approach. It has been successfully applied for automated fault detection and diagnosis of induction machines (Shukri et al., 2004; Altug, Chow and Trussell, 1999). Recently, ANFIS and its combination with other methods were also employed as an enhanced tool for fault classification. Some examples of the combined algorithms are ANFIS with genetic algorithms (Lei et al., 2009) and ANFIS with wavelet transform (Lou and Loparo, 2004) for bearing fault diagnosis. ANFIS has been also applied to classify the faults of induction motor with variable driving speed (Ye, Sadeghian and Wu, 2006).

Generally, the data obtained from measurements are of high dimensionality and have a large amount of redundant features. If the data are directly input into the classifier, the performance will be significantly decreased. Feature extraction and selection have been utilized for reducing dimension of data by selecting important features, with feature extraction and transformation of existing features into a lower dimensional space (Yang, Han and Yin, 2006). Nevertheless, each feature set contains many redundant or irrelevant features, along with salient features in feature space after feature extraction has been done. Consequently, there is a need for a feature selection tool to achieve good learning, classification accuracy, compact and easily understood knowledge-base, and reduction in computational time (Kumar, Jayaraman and Kulkarni, 2005).

In this study, CART (Breiman et al., 1984) and ANFIS are utilized as feature selection tool and classifier for fault diagnosis of induction motors, respectively. The proposed system consists of two stages. First, CART is performed to obtain the valuable features and identify the structure of classifier in the next iterative step. Second, the ANFIS classifier is used to diagnose the faults of induction motors in which the parameters of membership functions are tuned throughout the learning process.

1.2. Machine condition prognosis

Prognosis is the ability to predict accurately future health states and failure modes based on current health assessment and historical trends (Byington et al., 2003). There are two main functions of machine prognosis: failure prediction and remaining useful life (RUL) estimation. Failure prediction, which is addressed in this paper, allows pending failures to be identified before they come to more serious failures, resulting in machine breakdown and repair costs. RUL is the time left before a particular fault occurs or a part needs to be replaced.

The techniques related to prognosis can be broadly classified as experience-based, model-based, and data-driven techniques.

Experience-based prognostic approaches require component failure history data or operational usage profile data. They involve collecting statistical information from a large number of component samples to indicate survival duration of a component before a failure occurs and use these statistical parameters to predict the RUL of individual components. Generally, they are the least complex forms of prognostic techniques and their accuracy is limited because they base solely on the analysis of past experience.

Model-based prognostic approaches are applicable to where the accurate mathematical models can be constructed based on the physical fundamentals of a system. These approaches use residuals as features, which are the outcomes of consistency checks between the sensed measurements of system and the outputs of a mathematical model (Fu et al., 2004). Some of the studies, using these approaches, are described in Fu et al. (2004), Abbas et al. (2007), Li et al. (1999), Li, Kurfess and Liang (2000), Watson et al. (2005) or Luo et al. (2005). However, 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. Changes in structural dynamics and operating conditions can affect the mathematical model, which cannot mimic all real-life situations.

Data-driven prognosis techniques utilize and require a large amount of historical failure data to build a prognostic model that learns the system behavior. Among these techniques, artificial intelligence is regularly used because of its flexibility in generating appropriate models. Schwabacher and Goebel (2007) give a survey on artificial intelligent techniques used in prognosis. Other examples of data-driven prognosis techniques can be found in Vachtsevanos and Wang (2001), Huang et al. (2007), Wang, Golnaraghi and Ismail (2004), Brown et al. (2007). In comparison with other prognosis techniques, data-driven prognosis techniques are the most promising and effective techniques in machine condition prognosis. They frequently use vibration signals for temporal pattern identifications since it is relatively easy to measure and record machine vibration data. Accordingly, data-driven prognosis technique with vibration-based measurement are developed and used for machine condition prognosis in this study.

In addition, the more precisely the future states are predicted, the more effective the maintenance activities become. For that reason, long-term prediction methodology is considered in machine condition prognosis as significant. Yet, forecasting the future with long-term prediction strategy is a difficult and challenging task in time series prediction domain due to the growing uncertainties arising from unrelated sources, such as accumulation errors and insufficient information (Ji et al., 2005). The techniques of long-term prediction methodology will be described in the next section.

In long-term prediction, embedding dimension (ED), time delay (TM), and selection prediction models are essential. ED and TM are used to reconstruct the state space of machine condition time series and establish the fundamental parameters of the prediction model. ED is the number of initial observations that should be used as the input to prediction model. This value can be determined by using the false nearest neighbor method (FNN) (Kennel, Brown and Abarbanel, 1992), or the Cao's method (Cao, 1997). Of the two suggested methods, the FNN method is commonly used. TD is the number of steps that can be predicted by the prediction model to obtain the optimum performance. It can be calculated by using methods such as auto-correlation (Broomhead, 1986), average displacement (Rosenstein, Collins and Luca, 1994), and auto-mutual information (AMI) (Fraser and Swinney, 1986). In this study, AMI is chosen to estimate the time delay. After determining the embedding dimension and time delay, CART and ANFIS are utilized as the prediction model for comparing forecasting ability in long-term prediction of the machine condition.

2. Background knowledge and proposed systems

2.1. Background knowledge

2.1.1. Long-term prediction strategies

In time series domain, 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. According to Sorjamaa et al. (2007), there are three strategies mainly used in long-term prediction. They are the recursive, direct, and DirRec strategies for creating prediction model. The detailed information on these strategies can be found in Tran, Yang and Tan (2009).

2.1.2. Time delay (TM) estimation

There are several methods that can be used to choose the TM. However, most of them are based on empirical concepts and it is not easy to identify, which of the methods is suitable for a particular task. In this paper, TM is dealt with the auto mutual information (AMI) method. Mutual information (MI) can be used to evaluate dependence among random variables. The MI between two variables, X and Y, is the amount of information obtained from X in the presence of Y and vice versa. In the time series prediction problem, if Y is the output and X is a subset of the input variables, MI between X and Y is a 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 is presented in Fraser and Swinney (1986), Sorjamaa and Lendasse (2007) and Sorjamaa et al. (2007). AMI estimates the degree, to which the time series $x(t+\tau)$ can, on average, 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 a time series has been used to determine the optimal TM.

2.1.3. 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. Assume a time-series $x_1, x_2, ..., x_N$ and vector $y_i(d)$, given in equation (2), in a delay coordinate embedding of the time series with time delay τ and embedding dimension d are given.

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

The observations x_i are projections of the system 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, one can differentiate between points, which are "true" or "false" neighbors on the orbit. For instance, in Fig. 1, points A, B, C and D belong to a curve. In unidimensional projection, point D appears to be the nearest neighbor of A. However, point D is no longer the nearest neighbor of point A in two dimensions. In the same way, points A and C are nearest neighbors in two dimensions, but they are no neighbors when viewed in three dimensions. In this case, points A, D, C are examples of "false" neighbors while points A and B are "true" 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 Lou and Loparo (2004), the nearest neighbor is determined by finding the vector which minimizes the Euclidean distance:

$$R_d = \|y_i(d) - y_i^r(d)\|. \quad (3)$$

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

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

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

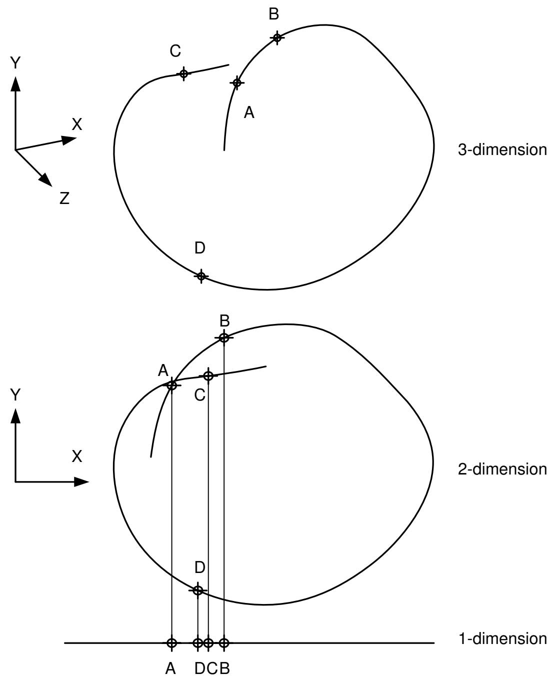


Figure 1. An example of false nearest neighbors

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 from practice. If both (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.

2.1.4. Classification and regression trees (CART) and adaptive neuro-fuzzy inference system (ANFIS) model

The CART algorithm was developed by Breiman et al. (1984) for classification or regression purposes depending on the response variable, which is either categorical or numerical. In case of classification, CART induces strictly binary trees through a process of binary recursive partitioning of feature space of a data set (Jang, Sun and Mizutani, 1996). Similarly, in case of regression, 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. The classification or regression trees are formed through two processes: tree growing and tree pruning. The detailed theory of these processes can be found in Breiman et al. (1984), Jang, Sun and Mizutani (1996) and Tran et al. (2008).

ANFIS is a fuzzy Sugeno model put in the framework of adaptive systems to facilitate learning and adaptation (Jang, 1993). Such framework makes the ANFIS modeling more systematic and less dependent on expert knowledge. In order to present ANFIS architecture, two fuzzy if-then rules based on a first-order Sugeno model are considered:

Rule 1: If (x is A_1) and (y is B_1) then $f_1 = p_1x + q_1y + r_1$.
Rule 2: If (x is A_2) and (y is B_2) then $f_2 = p_2x + q_2y + r_2$.

where x and y are the inputs, f_i are the outputs within the fuzzy region specified by the fuzzy rule, A_i and B_i are the fuzzy sets, $\{p_i, q_i, r_i\}$ is a set of design parameters, determined during the learning process. The ANFIS architecture to implement these rules and the learning algorithm to tune all the modifiable parameters can be found in Tran, Yang and Tan (2009), Jang, Sun and Mizutani (1996) and Ikonen and Najim (1996).

2.2. Proposed systems

2.2.1. Machine fault diagnosis

In this study, the vibration signals and current signals are utilized for detecting the faults of induction motors. The proposed system consists of four procedures as in Fig. 2: data acquisition, feature calculation, feature reduction, and fault classification. The summary role of each procedure is described below:

Data acquisition: this procedure is used to get the vibration and current signals. Furthermore, data processing is also carried out.

Feature calculation: the most significant features are calculated by using statistical feature parameters from time domain and frequency domain.

Feature selection: the CART algorithm is used to select the salient features from the whole feature set.

Fault classification: The data obtained from feature reduction procedure are split into two data sets: training and testing data. Training data are employed to build the model whilst testing data - to validate the model. The results indicate the accuracy of classification.

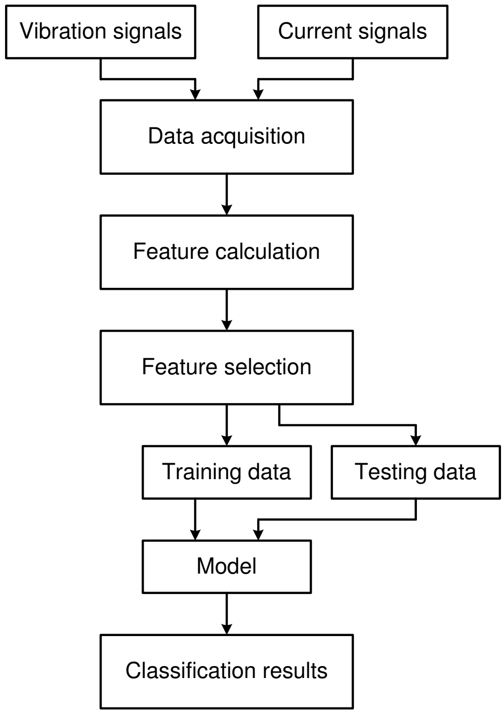


Figure 2. The proposed system for fault diagnosis

2.2.2. Machine condition prognosis

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

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.

Data splitting: the trending data obtained from previous procedure is split into two parts, the training and testing set. Different data are 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.

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.

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.

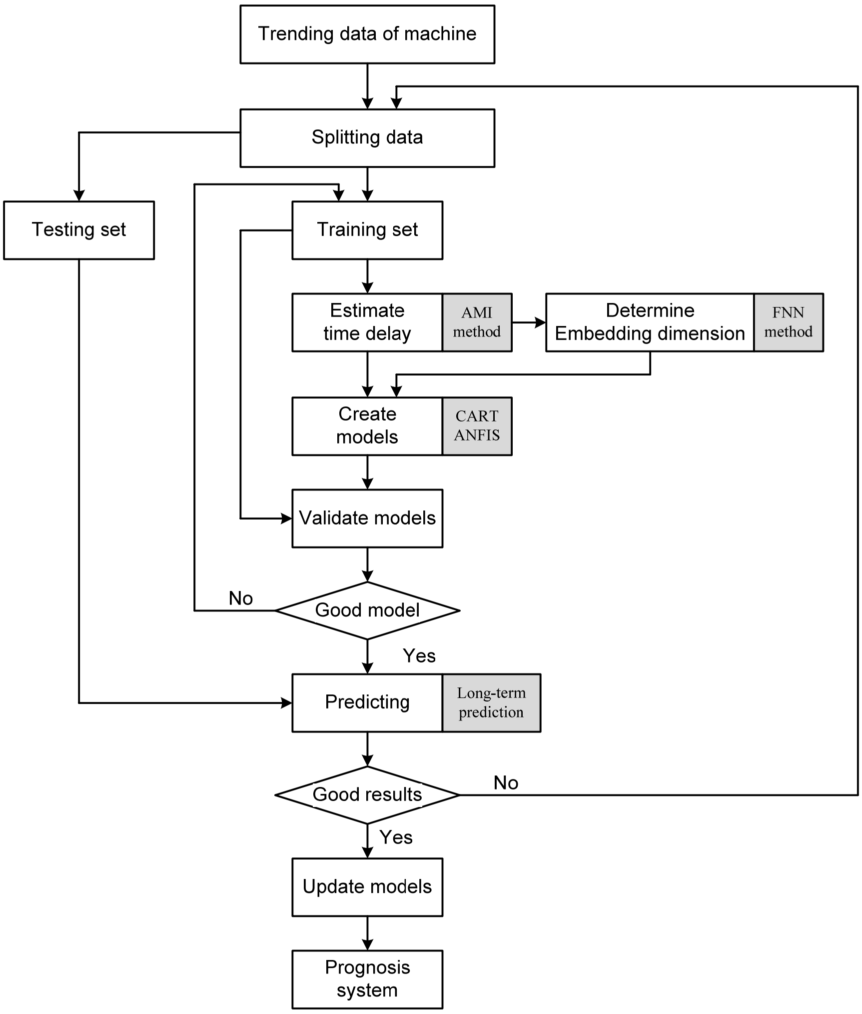


Figure 3. Proposed system for machine condition prognosis

3. Experiments

3.1. Machine fault diagnosis

To validate the CART-ANFIS model, experiment was carried out using a test-rig which consists of a motor, pulleys, belt, shaft and fan with changeable blade pitch angle that represents the load. The load can be changed by adjusting blade pitch angle or the number of blades. Six induction motors of 0.5 kW, 60 Hz, 4-pole were used to create data. One of the motors with good condition ("healthy") is used for comparison with faulty motors. The others are faulty motors, with rotor unbalance, broken rotor bar, phase unbalance, bearing outer race fault, bowed rotor, and adjustable eccentricity motor, as shown in Fig. 4. The conditions of faulty motors are described in Table 1.

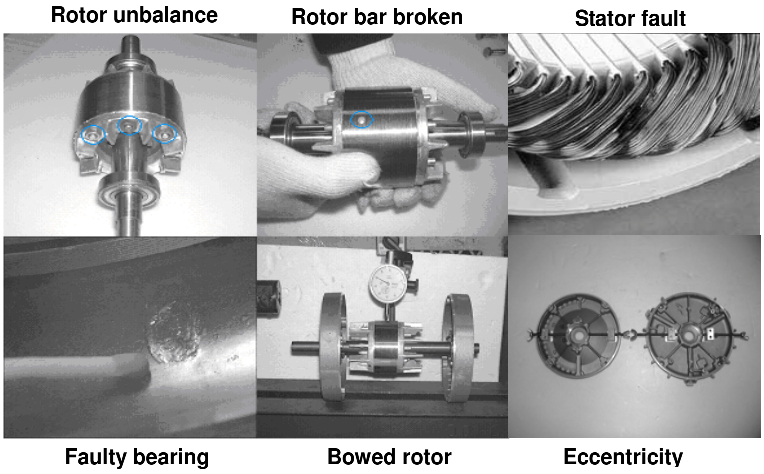


Figure 4. Faults on the induction motors

Table 1. Description of faulty motors

Fault condition	*Fault description*	*Other*
Broken rotor bar	Number of broken bar:12 ea	Total number of 34 bars
Bowed rotor	Max. shaft deflection: 0.075mm	Air-gap: 0.25mm
Faulty bearing	A spalling on outer raceway	#6203
Rotor unbalance	Unbalance mass on the rotor	8.4g
Eccentricity	Parallel and angular misalignments	Adjusting the bearing pedestal
Phase unbalance	Add resistance on one phase	8.4%

For acquiring data from the test rig, three AC current probes and three accelerometers were used to measure the stator current of the three-phase power supply and vibration signal in the horizontal, vertical and axial directions for evaluating the fault diagnosis system, respectively. The maximum frequency of the signal was 3 kHz, with 16,384 sampled data and giving a measured time of 2.133 seconds.

3.2. Machine condition prognosis

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

Table 2. Information on 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

The condition monitoring system of this compressor consists of off-line and on-line parts. In the off-line system, the vibration sensors were installed along axial, vertical and horizontal directions at the locations of drive-end motor, non drive-end motor, male rotor compressor and suction part of compressor. In the on-line system, acceleration sensors were located at the same places as in the off-line system, but only in the horizontal direction.

The trending data were recorded from August 2005 to November 2005, including 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 illustrated in Figs. 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.

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, indicating that possible faults were occurring in the machine. By disassembling and inspecting, these faults were identified as the damage of main bearings of the compressor (noted "Thrust: 7321 BDB") due to insufficient lubrication. Consequently, the surfaces of these bearings were overheated and delaminated (Tran et al., 2008). With the aim of forecasting the change of machine condition, the first 300 points will be used to train the system.

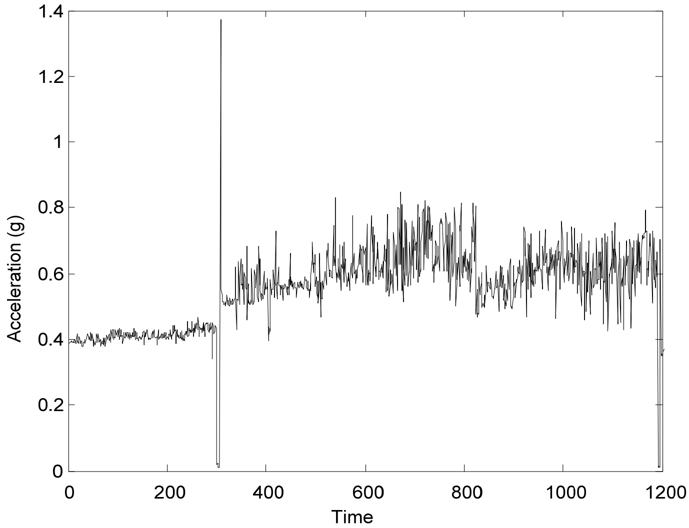


Figure 5. The entire peak acceleration data of low methane compressor

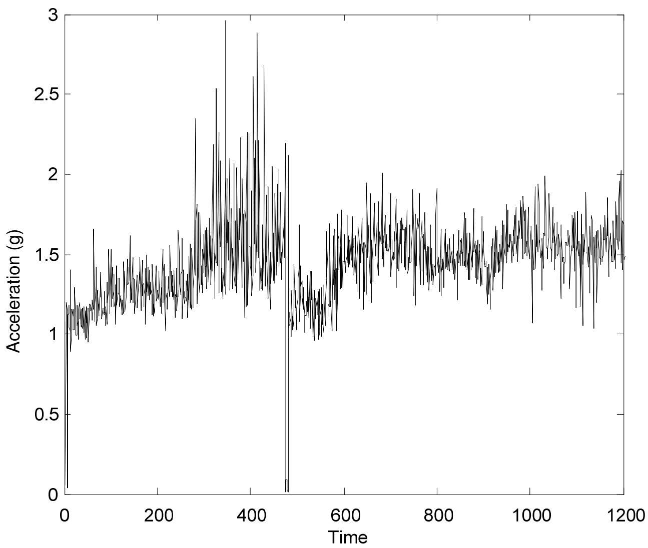


Figure 6. The entire envelope acceleration data of low methane compressor

4. Results and discussions

4.1. Machine fault diagnosis

4.1.1. Feature calculation

In this paper, feature calculation using statistical feature parameters from time and frequency domains was used. Sixty-three (21 parameters $\times$ 3 signals) features in total are calculated from 10 feature parameters of time domain. These parameters are mean, RMS, shape factor, skewness, kurtosis, crest factor, entropy error, entropy estimation and histogram of upper and lower limits. In addition, there are three parameters from frequency domain (RMS frequency, frequency center and root variance frequency), using the three direction vibration signals and three-phase current signals. The total number of feature parameters is shown in Table 3. The data sets of the features have 270 samples. In each operating condition, 20 samples are employed for training process and 10 samples for testing. The detailed descriptions of those data sets are shown in Table 4.

Table 3. Feature parameters

Signals	*Position*	*Feature parameters*	
		Time domain	*Frequency domain*
Vibration	Vertical	Mean	RMS variance frequency
	Horizontal	RMS	Frequency center
	Axial	Shape factor	Root variance frequency
Current	Phase A	Skewness	
	Phase B	Kurtosis	
	Phase C	Crest factor	
		Entropy error	
		Entropy estimation	
		Histogram lower	
		Histogram upper	

Table 4. Descriptions of data sets

Label of classification	*Condition*	*Number of training samples*	*Number of testing samples*
C1	Angular misalignment	20	10
C2	Bowed rotor	20	10
C3	Broken rotor bar	20	10
C4	Bearing outer race fault	20	10
C5	Mechanical unbalance	20	10
C6	Normal condition	20	10
C7	Parallel misalignment	20	10
C8	Phase unbalance ($30\circ$)	20	10
C9	Phase unbalance ($50\circ$)	20	10
Total samples		180	90

4.1.2. Feature selection and classification

The final trees, obtained from the feature sets corresponding to the vibration signals and current signals, are depicted in Figs. 7 and 8, respectively.

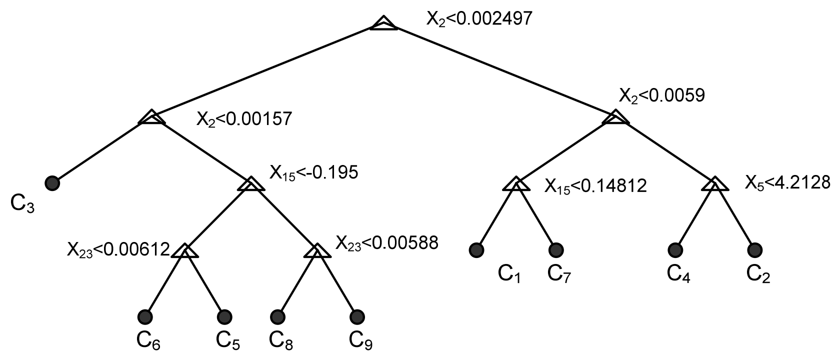


Figure 7. Tree of features obtained from vibration signals

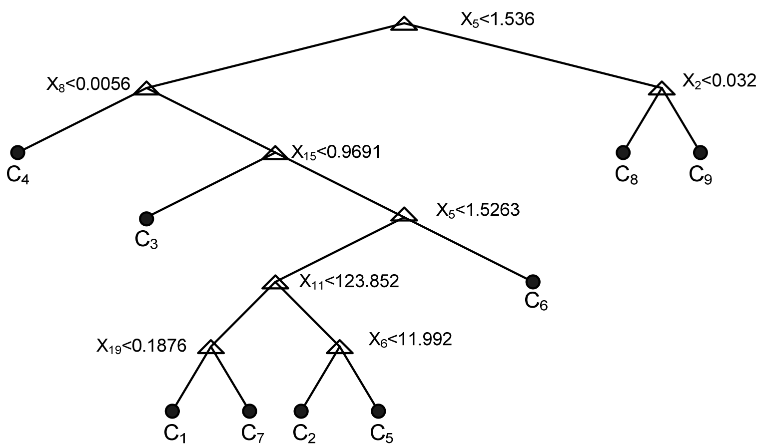


Figure 8. Tree of features obtained from current signals

Obviously, the feature appearing in root node of the trees is the most important one. The other features, in remaining nodes, appear in descending order of importance. It should be noted that only the features that contribute to the classification appear in the decision tree. Features, which have less discriminating capability, can be consciously discarded by deciding on the threshold. On this basis, the number of features is strongly diminished and only four features x_2,x_5,x_{15} and x_{23} of vibration signals and seven features ($x_2, x_5, x_6, x_8, x_{11}, x_{15}$

and x_{19}) of current signals are kept. Reduction of the number of features decreases the computational effort of ANFIS classifier in the next step. Furthermore, the trees are also used to identify the structure of ANFIS classifier. This structure includes the fuzzy rule set, which has been fuzzyfied from the crisp set of tree and membership functions, with bell-shaped functions being chosen for this purpose.

The system parameters and the chosen membership functions are automatically adjusted during the learning process. The convergence of the root mean squared (RMS) error is utilized to evaluate the learning process. If the rate of decrease of the RMS error and the performance are not significant, the learning process can be terminated. In this study, after 800 training epochs, the RMS error decreased to 0.087 and reached the convergent stage. This means that the learning process can be terminated.

The classification results are calculated using a ten-fold cross-validation evaluation where the data set to be evaluated is randomly partitioned, with 180 samples used for training and 90 samples for testing. The process is repeated with different random partitions and the results are averaged. The CART-ANFIS achieved 100% classification accuracy without any misclassification for 180 samples of training data for vibration and current signals. After training, the CART-ANFIS was tested against the testing data. The confusion matrix showing the classification results of the CART-ANFIS, established for 800 epochs of training cycle is given in Table 5. In this confusion matrix, each cell contains the number of samples that were correctly classified corresponding to actual network outputs and desired outputs of vibration signals and current signals. For example, the number shown as 10/7 in the first cell (first column and first row of the confusion matrix) means that there were 10 outputs belonging to class C1 and 7 outputs belonging to class C1 for vibration signals and current signals, respectively. This is similar for the other cells in the diagonal of confusion matrix. The cells off the diagonal of confusion matrix indicate the misclassifications. For example, the cell in first column and third row has the entry 0/1, showing that all subjects were correctly classified for vibration signals and one subject, which ought to belong to class C1, was classified as subject of class C3 for current signals.

The total classification accuracy for the test data was equal 91.11% with 8 misclassification out of 90 test samples for the vibration signal, and 76.67% with 21 misclassification for the current signal.

The test performance of the classifier can be evaluated with parameters such as sensitivity, specificity and total classification accuracy defined as:

Sensitivity: number of true positive decisions/ sum of number of true positive cases and number of false negative cases.

Specificity: number of true negative decisions/sum of number of true negative cases and number of false positive cases.

Total classification accuracy: number of correct decisions/total number of cases.

The values of statistical parameters are given in Table 6. The CART-ANFIS model classified subjects in classes C1 to C9, as shown in the form a/b, meaning the accuracy of classification corresponding to vibration signals and current signals, as follows: 100/70, 100/80, 70/90, 90/80, 80/70, 100/70, 90/70, 100/80 and 90/80%. These values are obtained from the diagonal cells of the confusion matrix. All of the data sets were classified with accuracy of 91.11%/76.67% (total classification accuracy).

Table 5. The confusion matrix for CART-ANFIS of 800 epochs

Output/	Confusion matrix (vibration/current signals)								
desired	C1	C2	C3	C4	C5	C6	C7	C8	C9
C1	10/7	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
C2	0/0	10/8	0/0	1/0	1/0	0/0	1/1	0/1	0/0
C3	0/1	0/0	7/9	0/1	0/1	0/2	0/0	0/0	1/2
C4	0/0	0/0	0/0	9/8	0/1	0/0	0/1	0/1	0/0
C5	0/0	0/0	0/0	0/0	8/7	0/0	0/0	0/0	0/0
C6	0/0	0/0	0/0	0/0	0/0	10/7	0/0	0/0	0/0
C7	0/0	0/0	0/0	0/0	0/0	0/1	9/7	0/0	0/0
C8	0/2	0/2	1/1	0/1	1/0	0/0	0/0	10/8	0/0
C9	0/0	0/0	2/0	0/0	0/1	0/0	0/1	0/0	9/8

Table 6. Values of statistical parameters

Datasets	Statistical parameters (vibration/current signals)		
label	Sensitivity (%)	Specificity (%)	Total classification accuracy (%)
C1	100/70	100/100	91.11/76.67
C2	100/80	96.5/97.5	
C3	70/90	98.75/91.25	
C4	90/80	100/96.25	
C5	80/70	100/100	
C6	100/70	100/100	
C7	90/70	100/98.75	
C8	100/80	97.5/92.5	
C9	90/80	97.5/97.5	

4.2. Machine condition prognosis

Before being used to generate the prediction models, TM is initially calculated according to the method mentioned in Section 2.1.2. Theoretically, the optimal time delay is the value at which the AMI reaches the first local minimum. From Fig. 9, the optimal TM of peak acceleration training data is found as equal 7. Similarly, 5 is the optimal TM value of envelope acceleration training data.

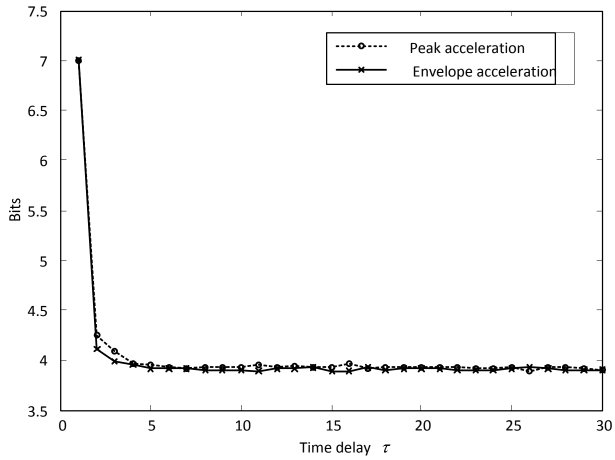


Figure 9. Time delay estimation

Using the FNN method, described in Section 2.1.3, the optimal TM is subsequently utilized to determine the embedding dimension d. For this, the tolerance level R_{tol} and threshold A_{tol} must be initially chosen. In the present study, R_{tol} = 15 and A_{tol} = 2 are used according to the results from Kennel, Brown and Abarbanel (1992). The relationship between the false nearest neighbor percentage and the embedding dimension for both peak acceleration data and envelope data is shown in Fig. 10. From this figure, the embedding dimension d is chosen as 4 for both data sets, since the false nearest neighbor percentage reaches then 0.

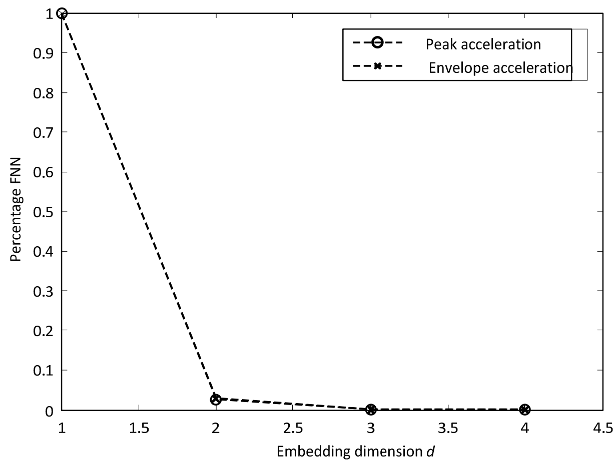


Figure 10. The relationship between FNN percentage and embedding dimension

Following the determination of TM and ED, the process of generating the prediction models is carried out. Based on those values, the training data are created, in which the number of observations is equal to ED and the number of predicted steps is equal to TM. Using this training data, the CART model and the ANFIS model are established. In case of the CART model, the number of response values for each terminal node in tree growing process is 5, and 10 cross-validations are used for selecting the best tree in tree pruning. Furthermore, in order to evaluate the predicting performance, the root-mean square error (RMSE) is utilized, i.e.

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

where N, y_i, 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.

Figs. 11(a) and 12(a) show the training and validating results of the CART models for peak acceleration and envelope acceleration data, respectively. The actual and predicted values are almost identical with very small RMSE, ranging from 0.002217 to 1.3314×10-5. This indicates that the learning capability of CART model is extremely good. Similarly, the ANFIS models are also created for both training sets of peak and envelope acceleration. There are four inputs for each ANFIS model due to the embedding dimension value. For each input, a bell shape is chosen for each membership function (MF) and the number of MFs is 2. This means that the region of values of each input is divided in two, namely, small and large. In order to evaluate the learning process, convergence of RMSE is also utilized. In this study, after executing 100 epochs, all RMSEs of the outputs reach the convergent stage for both the peak and the envelope acceleration data as shown in Fig. 13. Alternatively, the parameters of MFs are automatically adjusted through learning in order for the outputs of ANFIS model to match the actual values in training data. The changes of MF shapes are depicted in Fig. 14. The training and validating results of ANFIS models for both the peak and the envelope acceleration data are shown, respectively, in Figs. 11(b) and 12(b). From these figures, the RMSE values are obtained, equal 0.00876 and 0.08886. These values are slightly higher than those of the CART models. The reason could be that the number of MFs is improperly chosen. For higher accuracy of RMSEs, the number of MFs can be increased. Nevertheless, this will also increase the computational complexity and take too much training time.

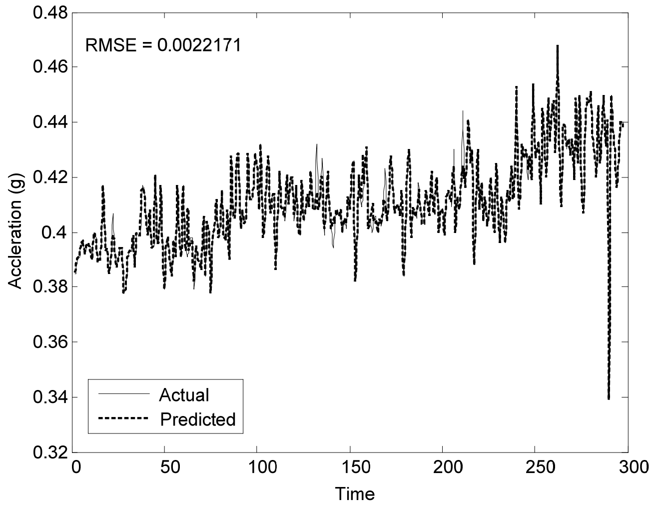
Figs. 15 and 16 show the prediction results of the CART and ANFIS models for peak acceleration and envelope acceleration data. The RMSE values of the CART and the ANFIS model for those data are summarized in Table 7. Although the RMSEs of ANFIS models are slightly higher than those of CART models in both cases of peak acceleration and envelope acceleration data, the predicted results of ANFIS models can keep track of the changes in the operating condition of machine more precisely. This is of crucial importance in industrial application of estimating the time-to-failure of equipments. As mentioned above, the prediction results of ANFIS models can be improved by adjusting the parameters of ANFIS. However, these changes should take into consideration the increase of computational complexity and of calculation time in the training process, which may lead to unrealistic characteristics for real life applications.

Table 7. The RMSEs of CART and ANFIS

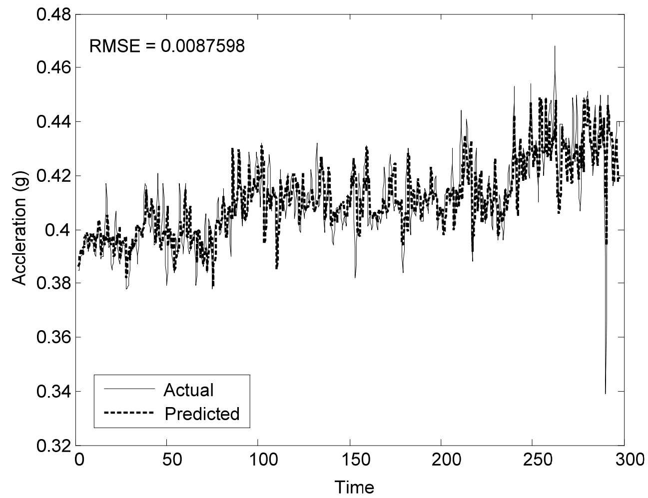
Data type	Training		Testing	
	CART	ANFIS	CART	ANFIS
Peak acceleration	0.002217	0.00876	0.14809	0.1708
Envelop acceleration	1.3314×10^{-5}	0.08886	0.2772	0.2938

5. Conclusion

Machine fault diagnosis and condition prognosis are essential in mechanical systems for detecting the faults and foretelling the degradation of operating conditions. In this study, an approach to machine fault diagnosis and condition prognosis based on CART and ANFIS has been investigated. In case of diagnosis, combined CART and ANFIS were used to perform fault diagnosis of induction motors. The classification results and statistical measures were used to evaluate the CART-ANFIS model. Total classification accuracy was 91.11% and 76.67% for vibration and current signals, respectively. In case of prognosis, long-term direct prediction for the operating conditions of machine based on data-driven approach was examined. The CART and ANFIS models were validated by the ability to predict future state conditions of a low methane compressor using peak acceleration and envelope acceleration data. Predictions of CART models are slightly better than those of ANFIS. Nonetheless, they are incapable of tracking the change of machines' operating conditions with as high accuracy as the ANFIS models. The capability of change-tracking of operating conditions is of crucial importance in estimating the RUL of industrial equipments. The results confirm that the proposed systems in both cases offer a potential for machine fault diagnosis and condition prognosis.

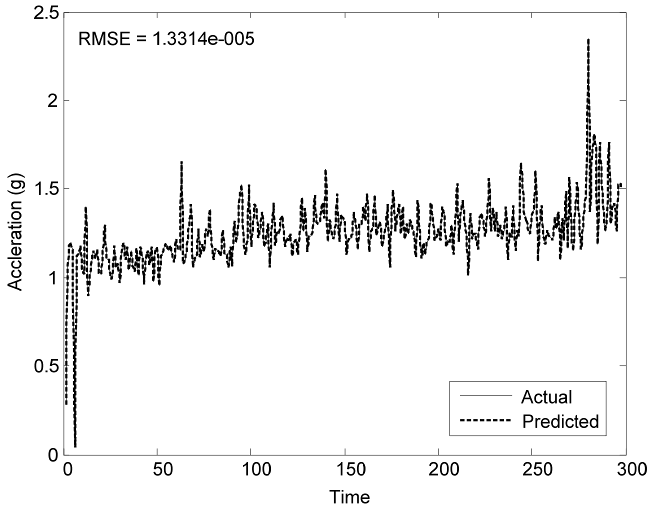


(a)

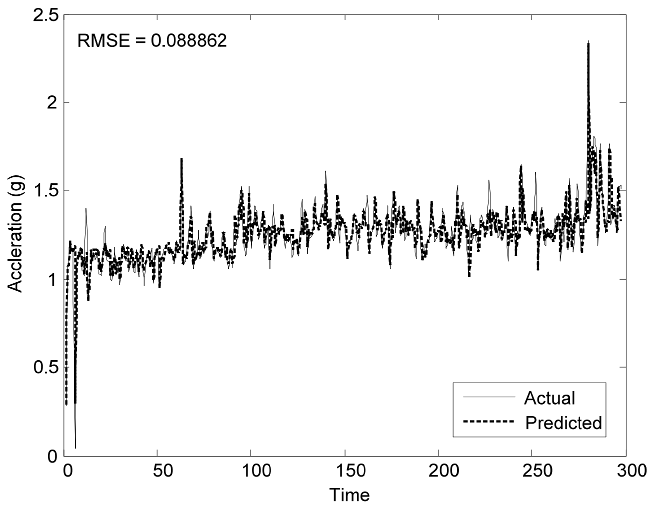


(b)

Figure 11. Training and validating results of peak acceleration data: (a) CART, (b) ANFIS

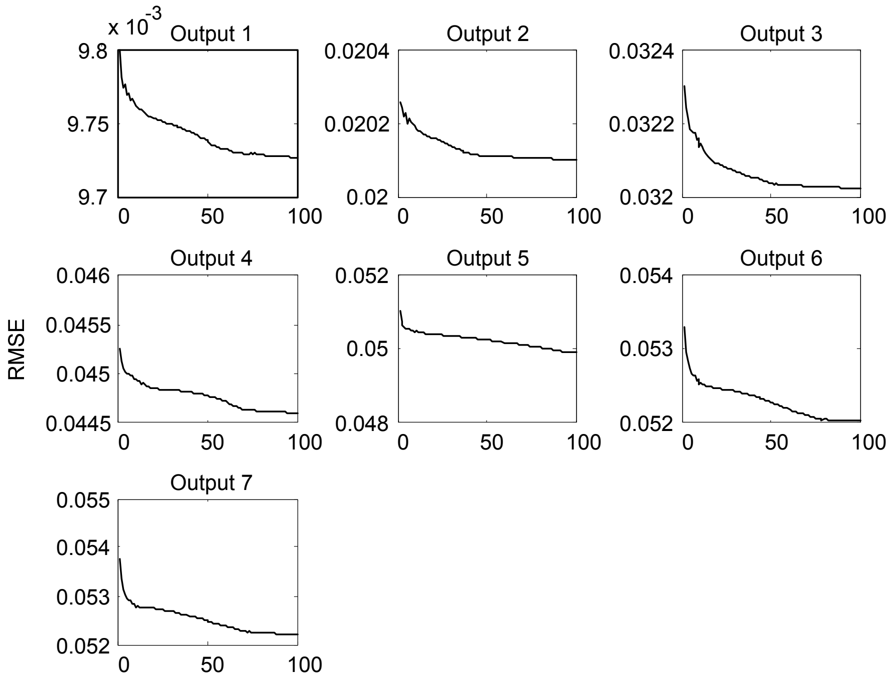


(a)

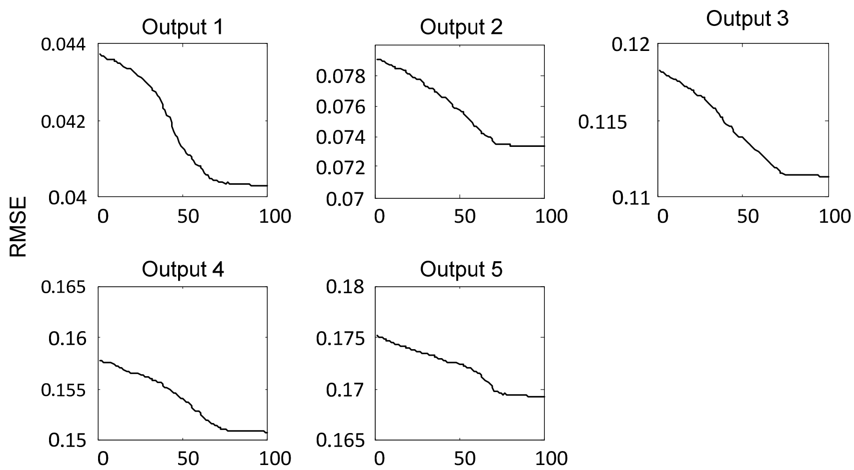


(b)

Figure 12. Training and validating results of envelope acceleration data: (a) CART, (b) ANFIS



(a)



(b)

Figure 13. The RMSE convergence curves: (a) Peak acceleration, (b) Envelope acceleration

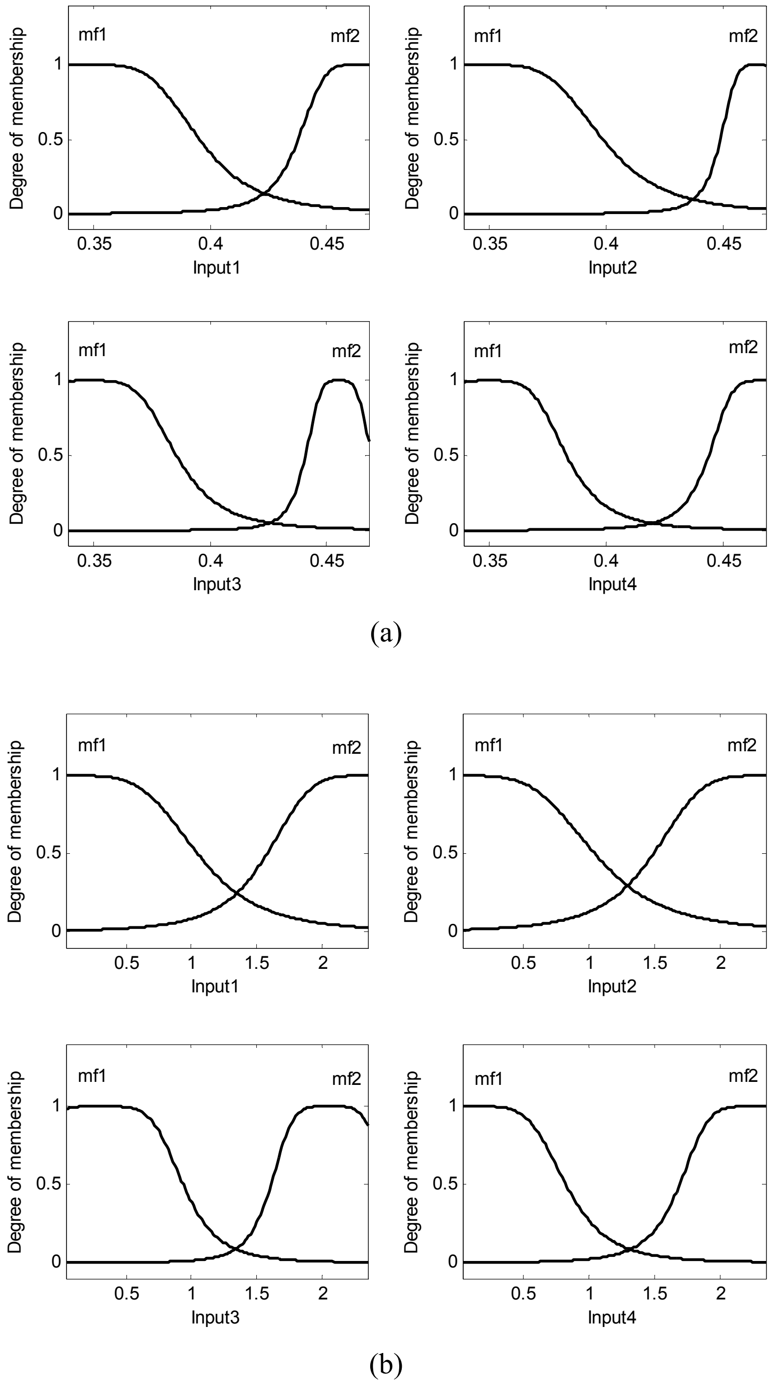
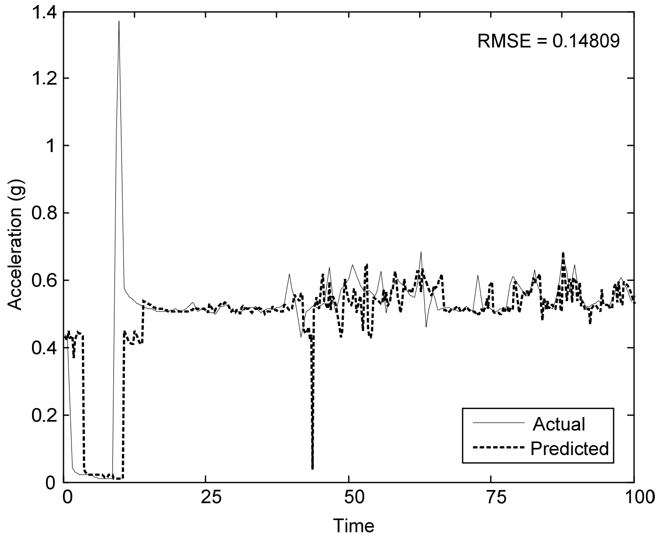
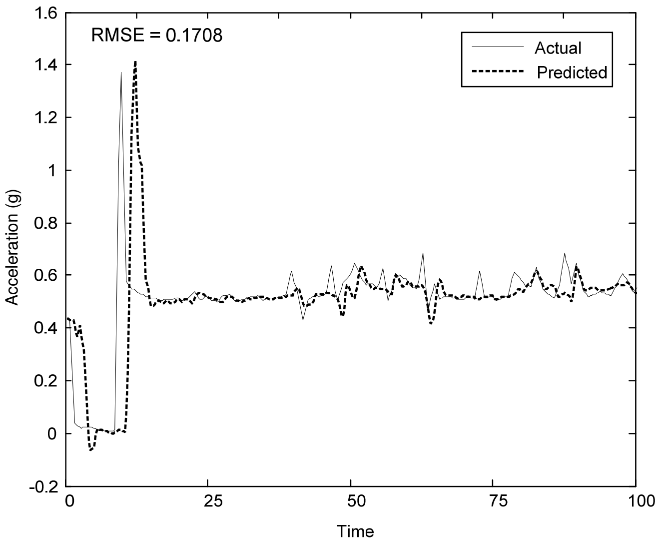


Figure 14. The changes of MFs after learning: (a) Peak acceleration, (b) Envelope acceleration

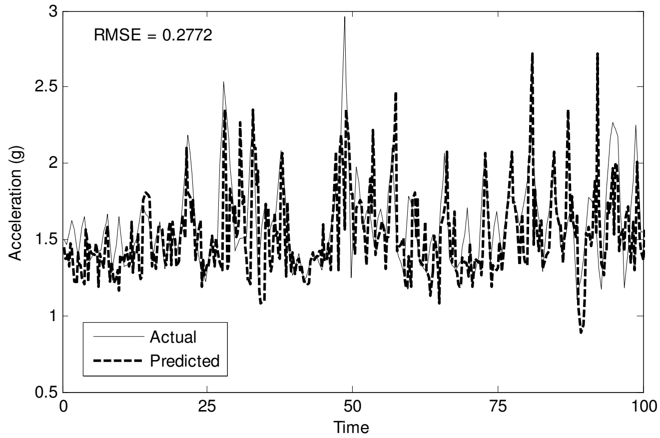


(a)

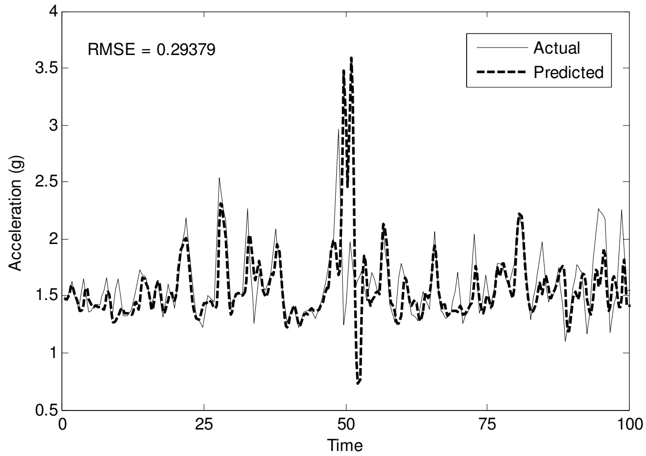


(b)

Figure 15. Prediction results for peak acceleration data: (a) CART, (b) ANFIS



(a)



(b)

Figure 16. Prediction results for envelope acceleration data: (a) CART, (b) ANFIS

References

ABBAS, M., FERRI, A.A., ORCHARD, M.E. and VACHTSEVANOS, G.J. (2007) An intelligent diagnostic/prognostic framework for automotive electrical systems. *Proc. 2007 IEEE Intelligent Vehicles Symposium*, Istanbul, 352-357.

ACOSTA, G.G., VERUCCHI, C.J. and GELSO, E.R. (2006) A current monitoring system for diagnosis electrical failures in induction motors. *Mechanical Systems and Signal Processing* **20**, 953-965.

ALTUG, S., CHOW, M.Y. and TRUSSELL, H.J. (1999) Fuzzy inference systems implemented on neural architectures for motor fault detection and diagnosis. *IEEE Trans. Industrial Electronics* **46** (6) 1069-1079.

BENBOUZID, M.E.H and NEJJARI, H. (2001) A simple fuzzy logic approach for induction motors stator condition monitoring. *Proc. IEEE IEMDC 2001*, 634-639.

BREIMAN, L., FRIEDMAN, J.H., OLSHEN, R.A. and STONE, C.J. (1984) *Classification and Regression Trees.* Chapman & Hall Press.

BROOMHEAD, D.S. (1986) Extracting qualitative dynamics from experimental data. *Physica* **D 20**, 217-236.

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. *Proc. of IEEE Aerospace Conference*, 1-12.

BYINGTON, C.S., WATSON, M., ROEMER, M.J., GALIC, T.R. and MCGROARTY, J.J. (2003) Prognostic enhancements to gas turbine diagnostic systems. *Proc. IEEE Aerospace Conference* 7, 3247-3255.

CAO, L. (1997) Practical method for determining the minimum embedding dimension of a scalar time series. *Physica* **D 110**, 43-50.

CASIMIR, R., BOUTLEUX, E., CLERC, G. and YAHOUI, A. (2006) The use of feature selection and nearest neighbors rule for faults diagnosis in induction motors. *Engineering Applications of Artificial Intelligence* **19**, 169-177.

CHEN, Y. (1995) A fuzzy decision system for fault classification under high levels of uncertainty. *J. of Dynamic Systems, Measurement, and Control* **117**, 108-115.

CHO, K.R., LANG, J.H. and UMANS, S. (1992) Detection of broken rotor bars using state and parameter estimation. *IEEE Trans. on Industry Applications* **28**, 702-713.

FRASER, A.M. and SWINNEY, H.L. (1986) Independent coordinates for strange attractors from mutual information. *Phys. Rev.* **A 33**, 1134-1140.

GOODE, P. and CHOW, M.Y. (1995) Using a neural/fuzzy to extract knowledge of incipient fault in induction motor: Part 1 - methodology. *IEEE Trans. Industrial Electronics* **42** (2) 131-138.

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.

Ikonen, E. and Najim, K. (1996) Fuzzy neural networks and application to the FBC process. *Proc. Control Theory Application* **143**, 259-269.

Isermann, R. (1984) Process fault detection based on modeling and estimation methods – a survey. *Automatica* **20**, 387-404.

Isermann, R. and Freyermuth, B. (1991) Process fault diagnosis based on process model knowledge – part I. *J. of Dynamic Systems, Measurement, and Control* **113**, 620-626.

Jang, J.S.R. (1993) ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans. System, Man and Cybernetics* **23** (3) 665-685.

Jang, J.S.R., Sun, C.T. and Mizutani, E. (1996) *Neuro-fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Prentice Hall.

Ji, Y., Hao, J., Reyhani, N. and Lendasse, A. (2005) Direct and recursive prediction of time series using mutual information selection. **LNCS 3512**, Springer, 1010-1017.

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.

Kumar, R., Jayaraman, V.K. and Kulkarni, R.D. (2005) An SVM classifier incorporating simultaneous noise reduction and feature selection: illustrative case examples. *Pattern Recognition* **38**, 41-49.

Lei, Y., He, Z., Zi, Y. and Hu, Q. (2009) Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs. *Mechanical Systems and Signal Processing*, in print.

Li, Y., Billington, S., Zhang, C., Kurfess, T., Danyluk, S. and Liang, S. (1999) Adaptive prognostics for rolling element bearing condition. *Mechanical Systems and Signal Processing* **13** (1) 103-113.

Li, Y., Kurfess, T.R. and Liang, S.Y. (2000) Stochastic prognostics for rolling element bearings. *Mechanical Systems and Signal Processing* **14** (5) 747-762.

Lou, X. and Loparo, K.A. (2004) Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mechanical Systems and Signal Processing* **18**, 1077-1095.

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. *Proc. 31st Annual Conference of IEEE Industrial Electronics Society*, New York, 2267-2272.

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.

Satish, B. and Samar, N.D.R. (2005) A fuzzy approach for diagnosis and prognosis of bearing faults in induction motors. *IEEE Power Engineering*

Society General Meeting **3**, 2291-2294.
SCHWABACHER, M. and GOEBEL, K. (2007) A survey of artificial intelligence for prognostics. *Proc. AAAI Fall Symposium on Artificial Intelligence for Prognostics*, 9-11 Nov. 2007.
SHUKRI, M., KHALID, M., YUSUF, R. and SHAFAWI, M. (2004) Induction machine diagnostic using adaptive neuron fuzzy inference system. In: M. Gh. Negoita et al., eds., *KES2004*, 380-387.
SOOD, A.K., FAHS, A.A. and HENEIN, N.A. (1985) Engine fault analysis – part I: statistical methods; part II: Parameter estimation approach. *IEEE Trans. on Industrial Electronics* **32**, 294-300 and 301-307.
SORJAMAA, A., HAO, J., REYHANI, N., JI, Y. and LENDASSE, A. (2007) Methodology for long-term prediction of time series. *Neurocomputing* **70**, 2861-2869.
SORJAMAA, A. and LENDASSE, A. (2007) Time series prediction as a problem of missing values: application to ESTSP and NN3 competition benchmarks. *Proc. European Symposium on Time Series Prediction*, 165-174.
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.
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.
TU, F., GHOSHAL, S., LUO, J., BISWAS, G., MAHADEVAN, S., JAW, L. and NAVARRA, K. (2004) PHM integration with maintenance and inventory management systems. *IEEE Aerospace Conference Digest*, Motana, 1-12.
VACHTSEVANOS, G. and WANG, P. (2001) Fault prognosis using dynamic wavelet neural networks. *Proc. IEEE Systems Readiness Technology Conference*, 857-870.
WANG, W.Q., GOLNARAGHI, M.F. and ISMAIL, F. (2004) Prognosis of machine health condition using neuro-fuzzy system. *Mechanical Systems and Signal Processing* **18**, 813-831.
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 Trans.* **48** (2) 208-217.
WIDODO, A., YANG, B.S. and HAN, T. (2007) Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors. *Expert Systems with Applications* **32** (2) 299-312.
YANG, B.S., HAN, T. and YIN, Z.J. (2006) Fault diagnosis system of induction motors using feature extraction, feature selection and classification algorithm. *JSME Int. J. (C)* **49** (3) 734-741.
YANG, B.S. and KIM, K.J. (2006) Application of Dempster-Shafer theory in fault diagnosis of induction motors using vibration and current signals. *Mechanical Systems and Signal Processing* **20**, 403-420.

YANG, B.S., JEONG, S.K., OH, Y.M. and TAN, A.C.C. (2004) Case-based reasoning with Petri nets for induction motors fault diagnosis. *Expert Systems with Applications* **27** (2) 301-311.

YE, Z., SADEGHIAN, A. and WU, B. (2006) Mechanical fault diagnosis for induction motor with variable speed drives using adaptive neuro-fuzzy inference system. *Electric Power Systems Research* **76**, 742-752.