We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

A Hybrid Fuzzy System for Real-Time Machinery Health Condition Monitoring

Wilson Wang
Lakehead University
Canada

1. Introduction

Rotary machinery is widely used in various types of engineering systems ranging from simple electric fans to complex machinery systems such as aircraft. A reliable online condition monitoring system is very useful in industries both as a quality control scheme and as a maintenance tool. In quality control, the early detection of faulty components can prevent machinery performance degradation and malfunction. As a maintenance tool, machinery health condition monitoring enables the establishment of a maintenance program based on an early warning. This can be of great value in cases involving critical machines (e.g., airplanes, power turbines, and chemical engineering facilities), where an unexpected shutdown can have serious economic or environmental consequences.

Condition monitoring is an act of fault diagnosis by means of appropriate observations from different information carriers, such as temperature, acoustics, lubricant, or vibration. Vibration-based monitoring, however, is the most commonly used approach in industries because of its ease of measurement, which also will be used in this study.

Fault diagnosis is a sequential process involving two steps: representative feature extraction and pattern classification. Feature extraction is a mapping process from the measured signal space to the feature space. Representative features associated with the health condition of a machinery component (or subsystem) are extracted by using appropriate signal processing techniques. Pattern classification is the process of classifying the characteristic features into different categories. The classical approach, which is also widely used in industry, relies on human expertise to relate the vibration features to the faults. This method, however, is tedious and not always reliable when the extracted features are contaminated by noise. Furthermore, it is difficult for a diagnostician to deal with the contradicting symptoms if multiple features are used. The alternative is to use analytical tools (Li & Lee, 2005, Gusumano et al., 2002) and data-driven paradigms (Isermann, 1998). The latter will be utilized in this work because an accurate mathematical model is difficult to derive for a complex mechanical system, especially when it operates in noisy environments. Data-driven diagnostic classification can be performed by reasoning tools such as neural networks (Rish et al, 2005, Uluyol, 2006), fuzzy logic (Mansoori et al., 2007, Ishibuchi & Yamamoto, 2005), and neural fuzzy synergetic schemes (Wang, 2008, Uluyol et al., 2006).

Even though several techniques have been proposed in the literature for machinery condition monitoring, it still remains a challenge in implementing a diagnostic tool for real-

world monitoring applications because of the complexity of machinery structures and operating conditions. When a monitoring system is used in real-time industrial applications, the critical issue is its reliability. Unreasonably missed alarms (i.e., the monitoring system cannot pick up existing faults) and false alarms (i.e., the monitoring system triggers an alarm because of noise instead of real faults) will seriously mitigate its validity. To tackle these challenges, the objective of this research work is to develop a new technique, an integrated classifier, for real-time condition monitoring in, especially, gear transmission systems. In this novel classifier, the monitoring reliability is enhanced by integrating the information of the object's future states forecast by a multiple-step predictor; furthermore, the diagnostic scheme is adaptively trained by a novel recursive hybrid algorithm to improve its convergence and adaptive capability.

This chapter is organized as follows: Section 2 describes integrated classifier, whereas the multiple-step predictor and monitoring indices are described in Section 3. Section 4 discusses the hybrid online training algorithm. In Section 5, the viability of the proposed integrated classifier is verified by experimental tests corresponding to different gear conditions.

2. Diagnostic system

The diagnostic classifier is used to integrate the selected features obtained by implementing appropriate signal processing techniques. The purpose is to make a more positive assessment of the health condition of the mechanical component (or subsystem) of interest. The diagnostic reliability in this suggested classifier will be enhanced by implementing the future (multi-step-ahead) states of the object's conditions. The forecasting in this integrated classifier is performed for input variables so as to make it easier to track the error sources in diagnostic operations.
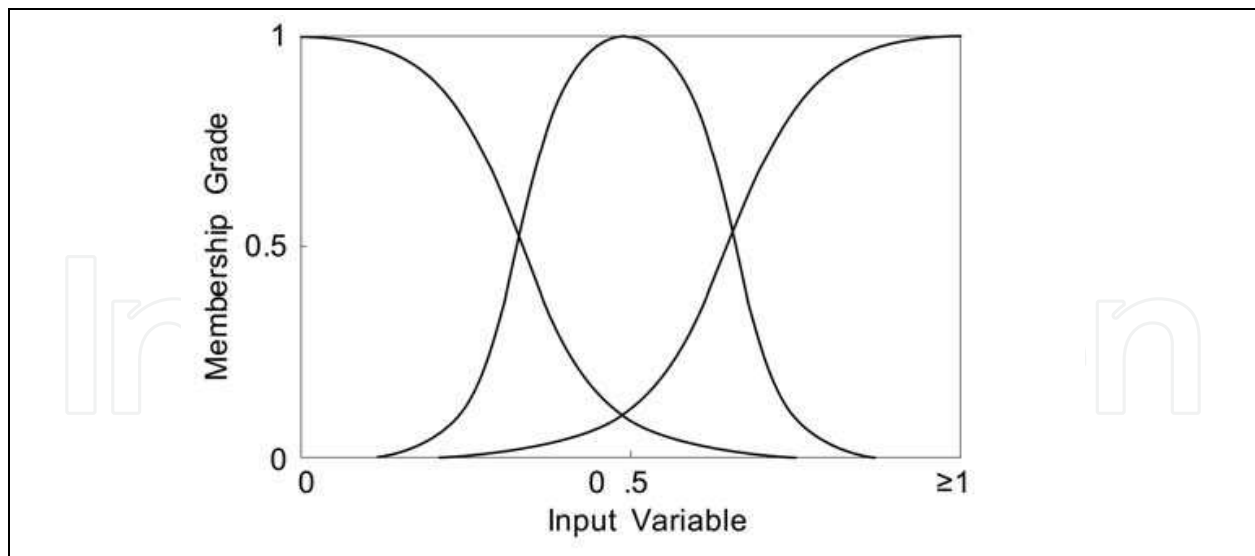

Fig. 1. The initial membership functions (MFs) for the input state variables.

The developed classifier is an NF paradigm which is able to facilitate the incorporation of diagnostic knowledge from expertise and to extract new knowledge in operations by online/offline training. The diagnostic classification is performed by fuzzy logic (Jang 1993), whereas an adaptive training algorithm, as discussed in Section 4, is utilized to fine-tune the fuzzy system parameters and structures. The conditions of each object (machinery

component or subsystem) are classified into three categories: *healthy* (C_1), *possible (initial) damage* (C_2), and *damage* (C_3), respectively. $\{x_1, x_2, …, x_n\}$ are the input variables at the current time step. Three membership functions (MFs), *small*, *medium*, and *large*, are assigned to each input variable with the initial states as shown in Fig. 1 where the fuzzy completeness (or the minimum fuzzy membership grade) is at 50%.

The diagnostic classification, in terms of the diagnostic indicator y, is formulated in the following form:

$$\Re_j : \text{If } (x_1 \text{ is } A_{1j}) \text{ and } (x_2 \text{ is } A_{2j}) \text{ and } \ldots \text{ and } (x_n \text{ is } A_{nj}) \Rightarrow (y \subset S_j \text{ with } w_j) \quad (1)$$

where A_{ij} are MFs; $i = 1, 2, …, n$, $j = 1, 2, …, m$, m denotes the number of rules; S_j represents one of the states C_1, C_2 or C_3, depending on the values of the diagnostic indicator.

When multiple features (input indices) are employed for diagnostic classification operations, the contribution of each feature combination (association) to the final decision depends, to a large degree, on the situation under which the diagnostic decision is made. Such a contribution is characterized by a weight factor w_j which is related to the feature association in each rule. The initial values of these rule weights are chosen to be unity; That is, all input state variables have initially assumed to have identical importance or robustness to the overall diagnostic output.

Similarly, the diagnostic classification based on the predicted monitoring indices, $\{ x'_1 , x'_2 , …, x'_n \}$, is formulated as:

$$\Re_j : \text{If } (x'_1 \text{ is } A_{1j}) \text{ and } (x'_2 \text{ is } A_{2j}) \text{ and } … \text{ and } (x'_n \text{ is } A_{nj}) \Rightarrow (y' \subset S_j \text{ with } w_j) \quad (2)$$

where y' is the diagnostic indicator based on forecast input variables.

The number of rules is associated with the diagnostic reasoning operations of input state variables. In general, if all monitoring indices are *small*, then the object is considered healthy (C_1). Otherwise, the object is possibly damaged. In this case, the diagnostic classification indicator y represents *faulty* condition only. Different feature association (rule) corresponds to a different confidence grade w_j in diagnosis. Fig. 2 schematically shows the network architecture of this integrated classifier. Unless specified, all the network links have unity weights.

The input nodes in layer 1 transmit the monitoring indices $\{x_1, x_2, …, x_n\}$ or their forecast future values $\{ x'_1 , x'_2 , …, x'_n \}$ to the next layer. These two sets of monitoring indices are input to the network and processed separately.

Each node in layer 2 acts as a MF, which can be either a single node that performs a simple activation function or multilayer nodes that perform a complex function. The nodes in layer 3 perform the fuzzy T-norm operations. If a product operator is used, the firing strength of rule $\Re_j$ is

$$\eta_j = \prod_{i=1}^{n} A_{ij}(x_i) \quad (3)$$

$$\eta'_j = \prod_{i=1}^{n} A_{ij}(x'_i) \quad (4)$$
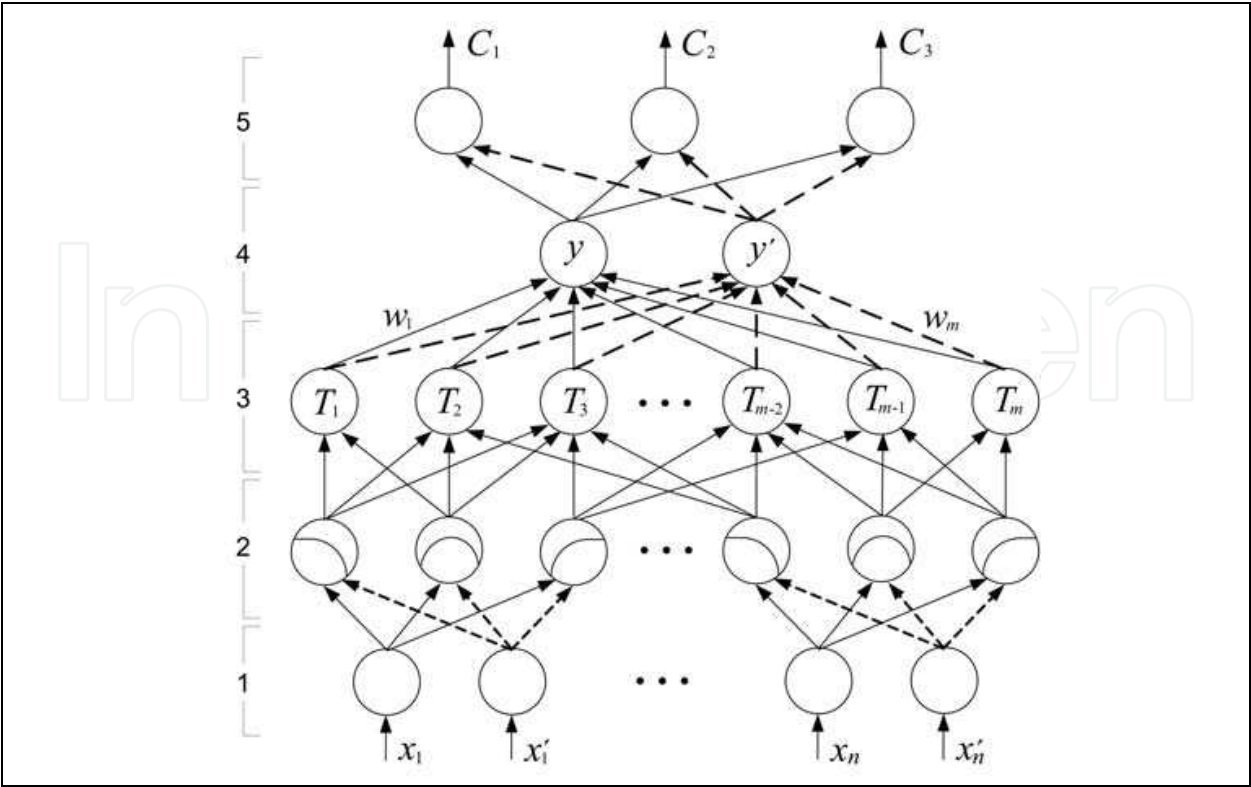
where $A_{ij}(\bullet)$ denote MF grades.

Fig. 2. The network architecture of the proposed integrated classifier.

Defuzzification is undertaken in layer 4. By normalization, the faulty diagnostic indicator will be

$$y = \frac{\sum_m \eta_j w_j}{\sum_m \eta_j} \tag{5}$$

Similarly, the fault diagnostic indicator based on forecast inputs will be

$$y' = \frac{\sum_m \eta'_j w_j}{\sum_m \eta'_j} \tag{6}$$

The states of the diagnostic indicator y (or y') are further classified into three categories:

$$\begin{cases} \text{If } 0 \leq y \leq 0.33 \;\Rightarrow\; Healthy\,(C_1) \\ \text{If } 0.33 < y \leq 0.66 \;\Rightarrow\; Possibly\ damaged\,(C_2) \,. \\ \text{If } 0.66 < y \leq 1 \;\Rightarrow\; Damaged\,(C_3) \end{cases}$$

The final decision regarding the health condition of the object of interest is made by:

a) If ($y \subset C_1$ and $y' \subset C_1$) or ($y \subset C_2$ and $y' \subset C_1$) then (the object is healthy C_1)

b) If ($y \subset C_3$ and $y' \subset C_3$) or ($y \subset C_2$ and $y' \subset C_3$) then (the object is damaged C_3) $\qquad$ (7)

c) Otherwise, (the object is possibly damaged C_2).

3. Prediction of monitoring indices

3.1 Monitoring indices

In general, most machinery defects are related to transmission systems, mainly for gears and bearings. In this work, gears are used as an example to illustrate how to apply the proposed integrated classifier for machinery condition monitoring. In operations, the fault diagnosis of a gear train is conducted gear by gear. Because the measured vibration is an overall signal contributed from various vibratory sources, the primary step is to differentiate the signal specific to each gear of interest by using a synchronous average filter (Wang et al., 2001). By this filtering process, the signals which are non-synchronous to the rotation of the gear of interest (e.g., those from bearings, shafts and other gears) are filtered out. As a result, each gear signal is computed and represented in one full revolution, called the *signal average* which will be used for advanced analysis by other signal processing techniques.

Several techniques have been proposed in the literature for gear fault detection. However, because of the complexity in the machinery structures and operating conditions, each fault detection technique has its own advantages and limitations, and is efficient for some specific application only (Wang et al., 2001). Consequently, the selected features for fault diagnostics should be robust, that is, sensitive to component defects but insensitive to noise (i.e., the signal not carrying information of interest). In this case, three features from the information domains of energy, amplitude, and phase are employed for the diagnosis operation:

1. Wavelet energy function, using the overall residual signal which is obtained by bandstop filtering out the gear mesh frequency $f_R N$ and its harmonics, where f_R is the rotation frequency (in Hz) of the gear of interest and N is the number of teeth of the gear;
2. Phase demodulation (McFadden, 1986), using the signal average;
3. Beta kurtosis, using the overall residual signal.

The details of these reference functions are listed in Appendix A.

Based on the derived reference functions, the monitoring indices are determined to quantify the feature characteristics. Each index is a function of two variables, magnitude and position. The magnitude of an index is determined as the normalized relative maximum amplitude value of the corresponding reference function; the position is where the maximum amplitude is located. Usually, the maximum amplitude positions in these reference functions do not coincide exactly due to the phase lags in signal processing. Based on simulation and test observations, an *influence window* is defined as a period of four tooth periods in this case. Correspondingly, if all indices are located within one influence window, one set of inputs $\{x_1, x_2, x_3\}$ is given to the classifier. Otherwise, if three indices are not within one influence window, the object has no fault or has more than one defect; more than one set of inputs should be provided to the classifier. For example, if x_3 does not fall within the influence window determined by x_1 and x_2, two sets of inputs will be given to the monitoring classifier: The first input vector is $\{x_1, x_2, x_3\}$, where x_3 is computed over the influence window determined by both x_1 and x_2; The second input vector is $\{x_1, x_2, x_3\}$, where x_1 and x_2 are determined over the influence window around x_3.

Fig. 3 illustrates an example of the reference functions corresponding to a healthy gear with 41 teeth. Fig. 3a shows part of the original vibration signal measured from the experimental setup to be illustrated in Section 5. Fig. 3b represents the signal average of the gear of interest, which is obtained by synchronous average filtering; each wave represents a tooth

period. Figs. 3c to 3e represent the resulting reference functions of the wavelet energy, beta kurtosis, and phase modulation, respectively. It is seen that no specific irregularities can be found from these reference functions for this healthy gear.
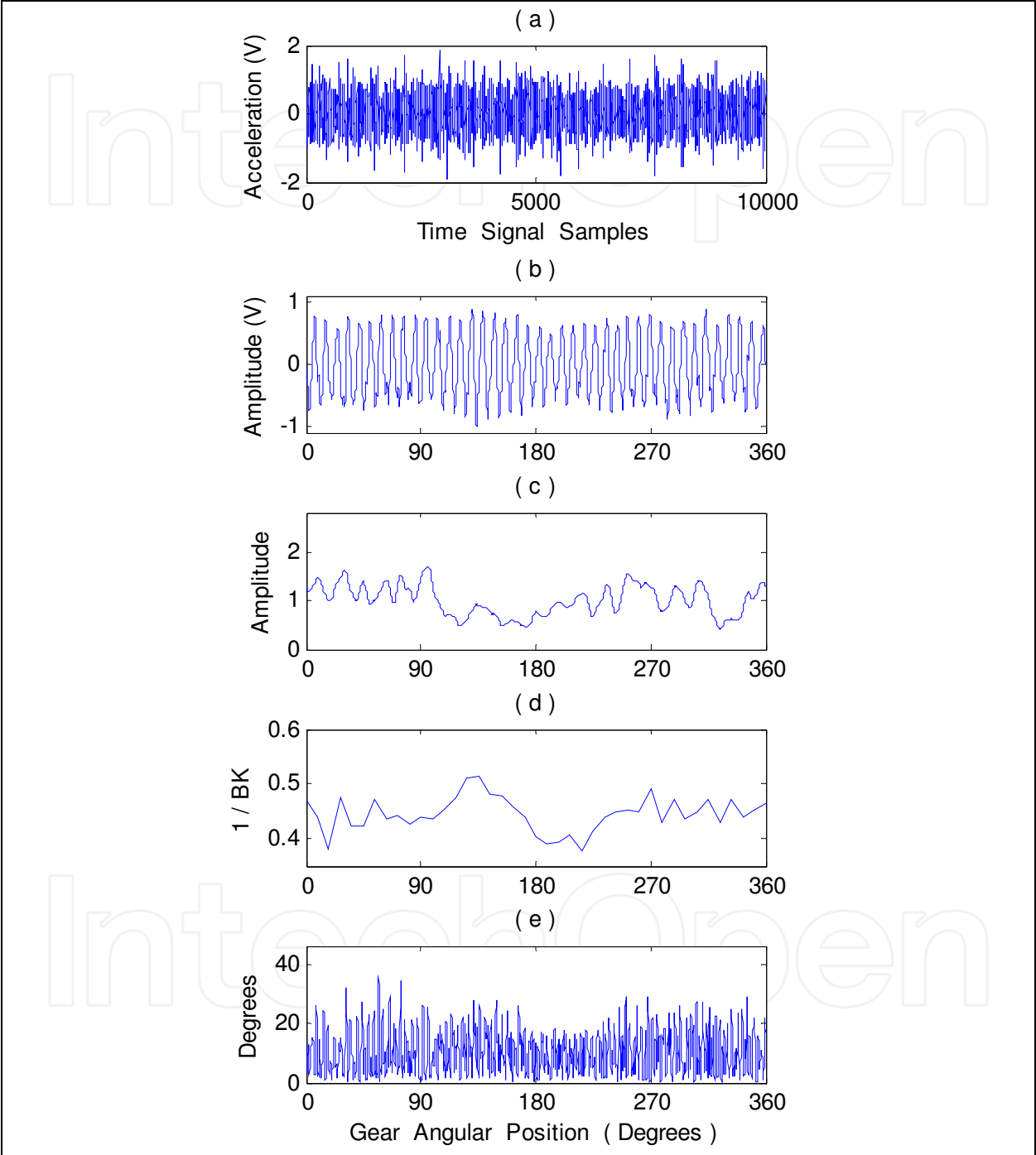

Fig. 3. Processing results for a healthy gear: (a) Part of the original vibration signal; (b) Signal average; (c) Wavelet reference function; (d) Beta kurtosis reference function; (e) Phase modulation reference function.

Fig. 4 shows the processing results corresponding to a cracked gear with 41 teeth. It is impossible to recognize the gear damage from the original signal (Fig. 4a). A little signature

irregularity can be recognized around 200° in the signal average graph (Fig. 4b). However, this gear damage can be identified clearly from the proposed reference functions (Figs. 4c to 4e). Although the maximum peak positions are little different from one graph to another, these peaks occur within one influence window (four tooth periods in this case).
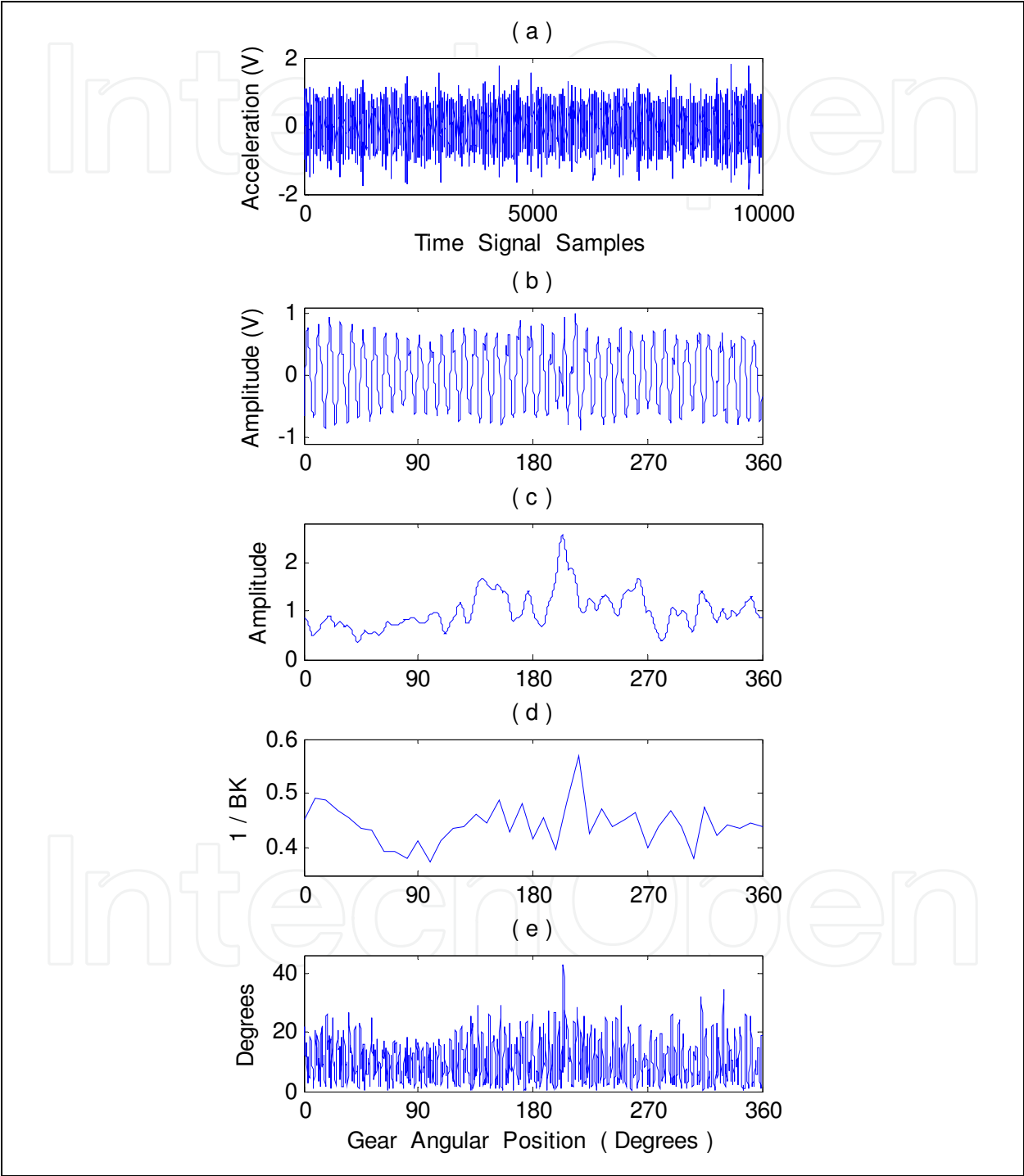

Fig. 4. Processing results for a cracked gear: (a) Part of the original vibration signal; (b) Signal average; (c) Wavelet reference function; (d) Beta kurtosis reference function; (e) Phase modulation reference function.

Fig. 5 illustrates the processing results for a chipped gear (with 41 teeth). Some signature irregularity can be recognized around 200° in the signal average graph (Fig. 5b) due to this gear tooth damage. However, this defect can be clearly identified from other three reference functions (Figs. 5c to 5e), and the monitoring indices are located within one influence window (four tooth periods).
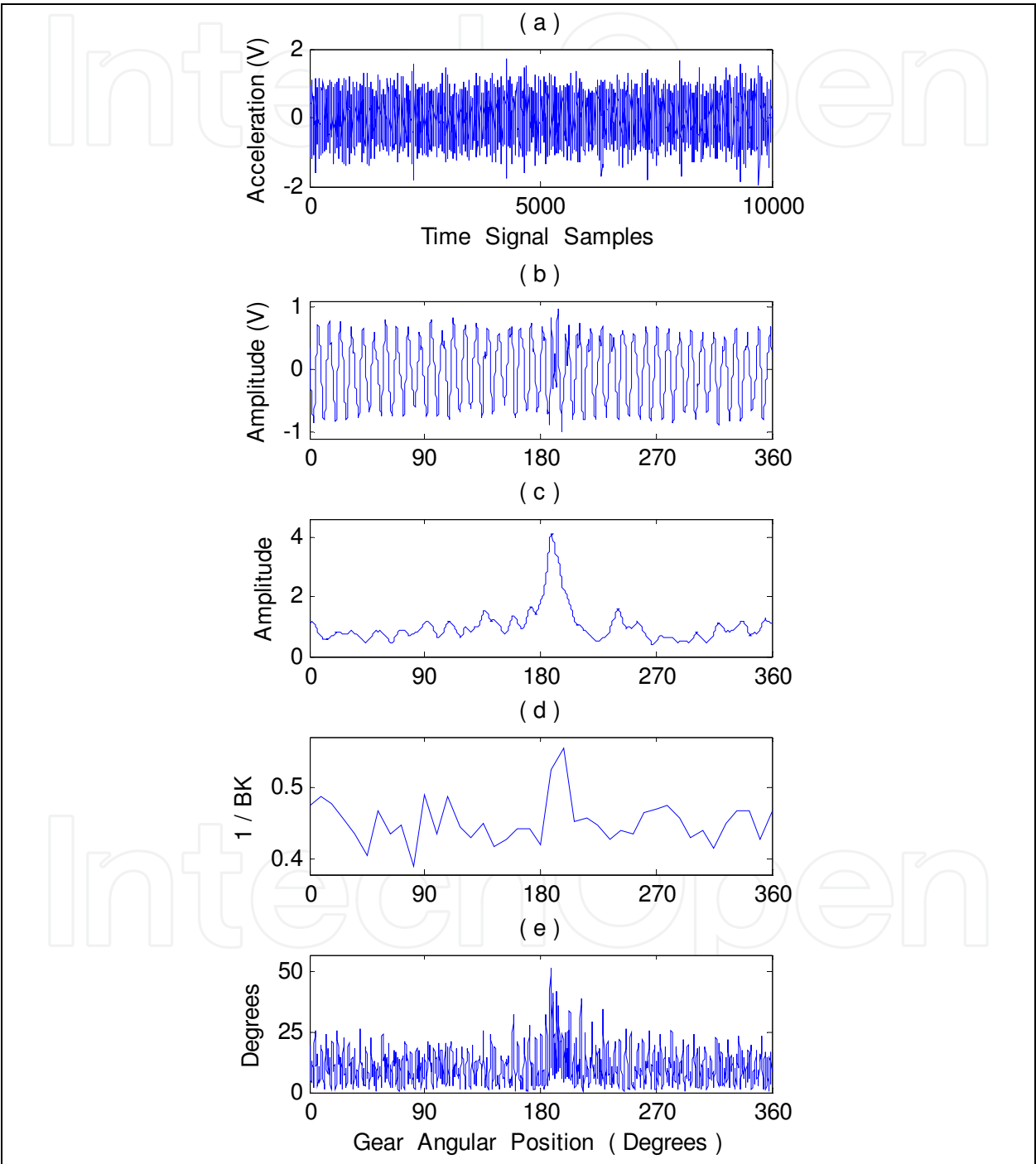

Fig. 5. Processing results for a chipped gear: (a) Part of the original vibration signal; (b) Signal average; (c) Wavelet reference function; (d) Beta kurtosis reference function; (e) Phase modulation reference function.

3.2 Forecasting of the monitoring indices

System state forecasting is the process to predict the future states in a dynamic system based on available observations. Several techniques have been suggested in the literature for time series forecasting. The classical methods are the use of stochastic models (Chelidze & Cusumano, 2004), which are usually difficult to derive for mechanical systems with complex structures. More recent research on time series forecasting has focused on the use of data-driven paradigms, such as neural networks and neural fuzzy schemes (Tse & Atherton, 1999, Pourahmadi, 2001). In this work, the multi-step-ahead prediction of the input variables (indices) is performed by the use of a predictor as suggested in (Wang & Vrbanek, 2007), whose effectiveness has been verified: it can capture and track the system's dynamic characteristics quickly and accurately, and it outperforms to other related classical forecasting schemes.

Given a monitoring index x_1, or x_2, or x_3, if $\{v_0 \ v_{-r} \ v_{-2r} \ v_{-3r}\}$ represent its current and previous three states with an interval of r steps, the r-step-ahead state v'_{+r} is estimated by a TS-1 fuzzy formulation:

$$\Re_j: \quad \text{If } (v_0 \text{ is } B_{0k}) \text{ and } (v_{-r} \text{ is } B_{1k}) \text{ and } (v_{-2r} \text{ is } B_{2k}) \text{ and } (v_{-3r} \text{ is } B_{3k})$$

$$\text{then } v'_{+r} = c_0^j v_0 + c_1^j v_{-r} + c_2^j v_{-2r} + c_3^j v_{-3r} + c_4^j \tag{8}$$

where $B_{\bullet}$ are MFs, c_i^j are constants, $i = 0, 1, ..., 3$; $j = 1, 2, ..., 16$; $k = 1, 2$. Fig. 6 illustrates its fuzzy reasoning architecture.
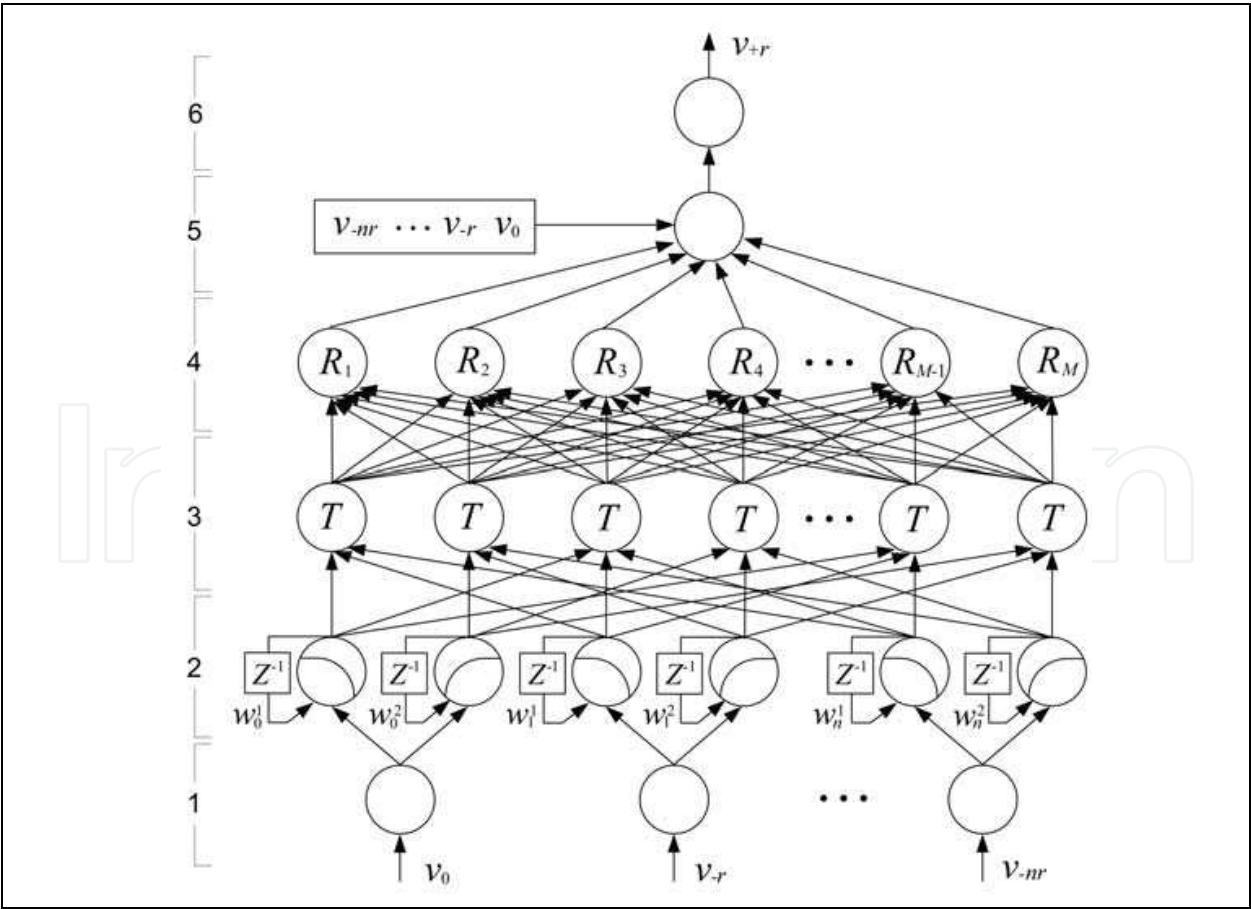

Fig. 6. The network architecture of the multi-step predictor.

This NF predictor has a weighted feedback link to each node in layer 2 to deal with time explicitly as opposed to representing temporal information spatially. The context units copy the activations of output nodes from the previous time step, and allow the network to memorize clues from the past, which forms a context for current processing. This function of recurrent networks is valuable for predictors with limited and step inputs (i.e., $r > 1$), to provide more information to the network so as to improve forecasting accuracy. If two sigmoid MFs are assigned to each input variable, the node output at the kth process step will be

$$\mu_{B_i^j}(V_{-ir}) = \frac{1}{1 + \exp[-a_i^j(V_{-ir} - b_i^j)]} \tag{9}$$

$$V_{-is} = v_{-ir}^{(k)} + w_i^m \mu_{B_i^j}(v_{-ir}^{(k-1)}) = v_{-is}^{(k)} + \frac{w_i^m}{1 + \exp[-a_i^j(v_{-ir}^{(k-1)} - b_i^j)]} \tag{10}$$

where m = 1, 2; i = 0, 1, . . ., n. $v_{-ir}^{(k)}$ and $v_{-ir}^{(k-1)}$ are, respectively, the input v_{-ir} at the kth and (k-1)th time steps, where k = 1, 2, . . ., K, K is the total number of time steps (or training data sets). If a max-product operator is applied in layer 3, and a centroid method is used for defuzzification in layer 5, by some related fuzzy operations, the predicted output v'_{+r} can be determined by

$$v'_{+r} = \sum_{j=1}^{16} \overline{\mu}_j \left(c_0^j v_0 + c_1^j v_{-r} + c_2^j v_{-2r} + c_3^j v_{-3r} + c_4^j\right) \tag{11}$$

where $\overline{\mu}_j = \dfrac{\mu_j}{\sum\limits_{j=1}^{16} \mu_j}$ denotes the normalized rule firing strength, and μ_j is the firing strength of the jth rule.

The fuzzy system parameters are trained by using a hybrid algorithm: that is, the premise parameters in the MFs $B_\bullet$ are trained by a real-time recurrent training algorithm whereas the consequent parameters c_i^j in (8) are updated by least squares estimate (LSE). Details about the training algorithm can be found in (Wang, 2008).

4. Online training of the diagnostic classifier

The developed diagnostic classifier should be optimized in order to achieve the desired input-output mapping. Several training algorithms have been proposed in the literature for NF-based classification schemes (Figueiredo et al., 2004, Castellano et al., 2004). In offline training, representative data should cover all of the possible application conditions (Korbicz et al., 2004); such a requirement is usually difficult to achieve in real-world machinery applications because most machinery operates in noisy and uncertain environments. Furthermore, machinery dynamic characteristics may change suddenly, for instance, just after repair or regular maintenance. Therefore, an adaptive training algorithm is preferred in time-varying systems to accommodate different machinery conditions (Wang & Lee, 2002). In this case, a hybrid method based on recursive Levenberg-Marquet (LM) and LSE will be

adopted to train the integrated classifier. Such a training approach possesses randomness that may help to escape certain local minima.

4.1 Training the premise MF parameters

The nonlinear premise MF parameters will be trained by adopting the recursive LM method. The general LM algorithm possesses quadratic convergence close to a minimum. Its convergence property is still reasonable, even if the initial estimates are poor. In addition, the LM algorithm has been proven globally convergent in many applications by properly choosing the step factors.

For a training data pair $\left\{x^{(p)}, d^{(p)}\right\}$, the inputs are $x^{(p)} = \left\{x_1^{(p)}, x_2^{(p)}, x_3^{(p)}\right\}$, $p = 1, 2, ..., P$; $d^{(p)}$ are the desired outputs {0, 0.5, 1} as $x^{(p)}$ belongs to C_1, C_2 and C_3, respectively. The error function with respect to adjustable MF parameters θ_p at the current time instant, p, is

$$E(\theta_p) = \frac{1}{2}\sum_{p=1}^{P}\left[y_p(\theta_p) - d_p\right]^2 = \frac{1}{2}\sum_{p=1}^{P}r_p^2(\theta_p) = r_p^T(\theta_p)r_p(\theta_p) \tag{12}$$

where $y_p(\theta_p)$ is the pth output determined by Eq. (5). $p = 1, 2, ..., P$; d_p is the desired output. To simplify expressions, the variable θ_p is dropped in the related terms in this section. r_p is the error vector that can be either linear or nonlinear. By taking the Taylor series expansion and neglecting higher order terms,

$$\theta_{p+1} \approx \theta_p + \lambda(J_p^T J_p + \eta I)^{-1}J_p^T r_p = \theta_p + (1-\alpha)H_p^{-1}J_p^T r_p \tag{13}$$

where $J_p \in \mathrm{R}^{N \times Z}$ denotes the Jacobian matrix; Z is the dimension (or the number of adjustable parameters) of θ_p; $H_p \in \mathrm{R}^{Z \times Z}$ is the modified Hessian matrix; $I \in \mathrm{R}^{Z \times Z}$ is an identity matrix; $\lambda = 1-\alpha$ is the learning rate, and α is the forgetting factor.

The Hessian matrix can be expressed as

$$H_p = \alpha H_{p-1} - (1-\alpha)(J_p^T J_p + \eta I) \tag{14}$$

In implementation, instead of computing the $Z \times Z$ matrix ηI at each time step, a diagonal element is added at each time step

$$H_p = \alpha H_{p-1} - (1-\alpha)(J_p^T J_p + Z\eta \Lambda) \tag{15}$$

where $\Lambda \in \mathrm{R}^{Z \times Z}$ has only one nonzero element located at $\{p \bmod(Z) + 1\}$ diagonal position:

$$\Lambda_{ii} = \begin{cases} 1, & if \ i = \{p \bmod(Z) + 1\} \\ 0, & otherwise \end{cases} \tag{16}$$

Correspondingly, (15) can be rewritten as

$$H_p = \alpha H_{p-1} - (1-\alpha)[UV^{-1}U^T] \tag{17}$$

where U is a $Z \times 2$ matrix whose first column is J_p and second column consists of a $Z \times 1$ vector with one element of 1 at the position of $\{p \bmod(Z) + 1\}$

$$U^T(\theta_p) = \begin{bmatrix} J_p^T \\ 0 \cdots 0 \; 1 \; 0 \cdots 0 \end{bmatrix}, \quad \text{and} \quad V^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & Z\eta \end{bmatrix}.$$

The computation of H_p^{-1} in (13) is time consuming, and is not suitable for real-time applications. To solve this problem, Eq. (13) is rewritten as

$$\theta_{p+1} = \theta_p + (1-\alpha)H_p^{-1}J_p^T r_p = \theta_p + (1-\alpha)\Big\{(\alpha H_{p-1})^{-1} - (\alpha H_{p-1})^{-1}(1-\alpha)U \times$$

$$\Big[V + U^T(\alpha H_{p-1})^{-1}(1-\alpha)U\Big]^{-1}U^T(\alpha H_{p-1})^{-1}\Big\}J_p^T r_p \tag{18}$$

Based on the matrix inversion formula and by some manipulations, Eq. (18) becomes

$$\theta_{p+1} = \theta_p + (1-\alpha)\Big\{\alpha H_{p-1} + (1-\alpha)U V^{-1} U^T\Big\}^{-1}J_p^T r_p \tag{19}$$

The recursive LM algorithm can be represented by

$$\theta_{p+1} = \theta_p + \Phi_p J_p r_p \tag{20}$$

$$\Phi_p = \frac{1}{\alpha}\left[\Phi_{p-1} - \frac{\Phi_{p-1}U U^T \Phi_{p-1}}{\alpha V + U^T \Phi_{p-1}U}\right] \tag{21}$$

The denominator $\alpha V + U^T \Phi_{p-1} U$ is a matrix with dimension 2×2; its inverse computation is simple, and can be implemented for real-time applications. $\theta_0 = 0$. Φ_p is a covariance matrix with initial condition $\Phi_0 = \rho I$, where ρ is a positive quantity and I is an identity matrix.

By simulation tests with the requirements of the recognition rate ≥ 80%, reasonable training speed and accuracy, the following initial values are given to the related parameters in this study: $\eta = 0.01$ with tested range of $\eta \in [0.001, 10]$; $\alpha = 0.995$ with tested range of $\alpha \in [0.95, 1]$; $\rho = 10^3$ with tested range of $\rho \in [10^2, 10^5]$.

4.2 Implementation of the hybrid training method

In implementation, inside each training epoch, the nonlinear MF parameters in the classifier are optimized in the backward pass by using a recursive LM method, whereas consequent linear rule weights are updated by LSE in the forward pass. On the other hand, after training or real applications over some time period, if the updated rule weights w_j are sufficiently small (e.g., $w_j < 0.01$), the contribution of the related rule to the final classification operation can be neglected, and that rule can be removed from the rule base.

5. Performance evaluation

5.1. Experimental setup

Fig. 7 shows the experimental setup used in this study to verify the performance of the proposed integrated classifier.

The apparatus is anchored onto a massive concrete block. It consists of a 3-HP AC drive motor and a gearbox. The motor rotation is controlled by a speed controller which allows tested gears operating in the range of 20 to 4200 rpm. An optical sensor provides a one-pulse-per-revolution signal which is used as the reference for the time synchronous average
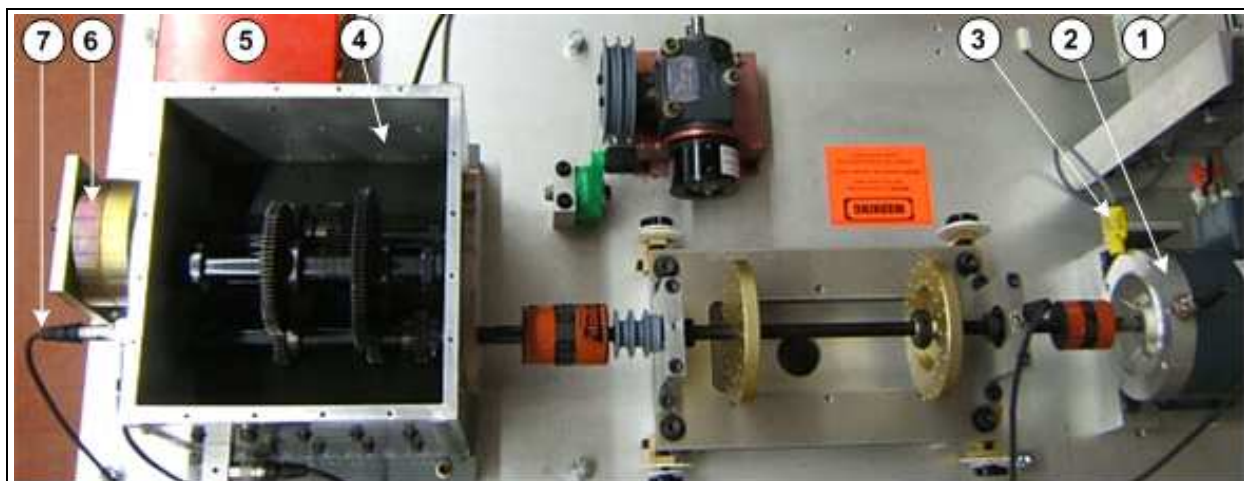
Fig. 7. The experimental setup: 1-speed controller, 2-motor, 3-optical sensor, 4-gearbox, 5-load controller, 6-loading system, 7-sensors.

filtering. The gearbox consists of two pairs of spur or helical gears. The shafts in the gearbox are mounted to the housing by rolling element bearings. The load is provided by a magnetic loading system which is connected to the output shaft. The speed of the drive motor and the load are adjusted to simulate different speed/load operating conditions. The vibration is measured using ICP accelerometers mounted on the gearbox housing along different orientations. After being properly preconditioned, the collected signals are fed to a computer for further processing.

5.2 Performance evaluation
To verify the viability of the proposed classifier, five gear cases are tested in this study as represented in Fig. 8:
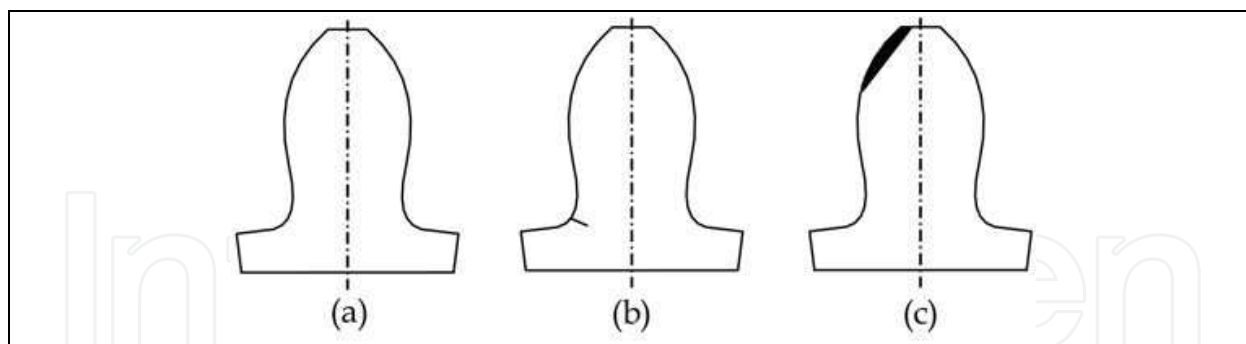

Fig. 8. Gear conditions tested: (a) healthy gear, (b) cracked gear; (c) chipped gear.

a.   healthy gears (C_1);
b.   gears having a tooth crack with 15% (C_2) and 50% (C_3) tooth root thickness;
c.   gears having a chipped tooth with 10% (C_2) and 40% (C_3) tooth surface area removed.

These demonstrated faults belong to localized gear defects. From the signal property standpoint, when a localized fault occurs, some high-amplitude pulses will be generated due to impacts, which are relatively easier for a signal processing technique to recognize. When a localized fault propagates towards a distributed defect, the overall energy of the fault will increase, but it often becomes more wideband in nature and difficult to detect in the presence of the other vibratory components of the machine. This example identifies a

characteristic of currently used fault detection techniques: It is usually easier to detect a distinct low-level narrowband tone than a high-level wideband signal in the presence of other signals or noises. Even though a distributed defect, such as pitting and wear, is initiated from a localized fault which is detectable as an incipient defect, most currently available vibration-based signal processing techniques cannot effectively detect an advanced distributed fault which, however, can be diagnosed based on other information carriers, such as acoustic signals.

To make a comparison, the diagnostic results from the following three classifiers are also listed:

1. A pure fuzzy system with a similar reasoning architecture as in Fig. 2 but without the use of predictors. The rule weight factors are chosen as those in the integrated classifier after initial training.
2. Classifier-1: An NF classifier with a similar reasoning architecture as in Fig. 2 but without predictors. Its MF parameters are trained by a gradient-LSE algorithm.
3. Classifier-2: Same as Classifier-1, but trained by the hybrid algorithm of the recursive LM and LSE.

Given the network architectures, the initial parameters of three adaptive classifiers can be primarily trained by using some data sets collected in previous tests on the same test apparatus, or be initialized by experience. Then these classifier parameters are optimized in the following online training processes.

During online tests, motor speed and load levels are randomly changed to simulate general and unusual machinery operating conditions. The tests are conducted under load levels from 0.5 to 3 hp, and motor speeds from 50 to 3600 rpm.

In online monitoring, based on test schedule and load/speed change frequency, the monitoring time-interval is set at 15 minutes; that is, all the monitoring schemes are applied automatically every 15 minutes for condition monitoring operations. Three-steps-ahead predictors (i.e., $r = 3$) are used in the integrated classifier. The selection of data size depends on noise reduction requirement; usually the data for the gear with the lowest speed should cover more than 100 revolutions. For example, if the slowest gear speed in the gearbox is 1200 rpm, the data acquisition process takes at least 5 seconds (15 seconds in this case). The monitoring is performed gear by gear. Three examples corresponding to healthy, cracked and chipped gears (all having 41 teeth) have been illustrated in Figs. 3 to 5, respectively.

Each healthy gear condition is tested over 24 hours whereas each faulty gear condition is tested over 50 hours. In total, 386 data pairs are recorded for testing purpose. Table 1 summarizes the classification performance by different diagnostic schemes.

Diagnostic Schemes	Healthy Gears		Cracked Gears		Chipped Gears		Overall Accuracy
	M.A.	F.A.	M.A.	F.A.	M.A.	F.A.	
Fuzzy System	0	13	12	16	3	8	85.3 %
Classifier-1	0	7	6	9	1	5	92.5 %
Classifier-2	0	5	7	8	0	4	93.6 %
New Classifier	0	2	3	3	0	1	97.6 %

Table 1. Comparison of the diagnostic results from different diagnostic schemes. M.A.-Missed Alarms, F.A.- False Alarms.

The fuzzy classifier records 15 missed alarms and 37 false alarms, with an overall reliability of 85.3%. Its relatively poor diagnostic performance is mainly due to the lack of learning capability. In addition, fixed or human-determined system parameters are subject to variations and are rarely optimal in terms of reproducing the desired classification outputs, which results in the fuzzy classifier not being optimized under different operating conditions.

Classifier-1 records 7 missed alarms and 21 false alarms, with an overall reliability of 92.5%. One difference between this NF system and the fuzzy classifier is related to the rule weight factors. Each signal processing technique (and the resulting feature) has a limited capability in fault detection. Even if the firing strengths of two fuzzy if-then rules are identical, their diagnostic reliabilities may be different under different machinery conditions. Therefore, rule weights play an important role in the diagnostic classification operations.

Classifier-2 records 7 missed alarms and 17 false alarms, with an overall reliability of 93.6%. The main difference between Classifier-2 and Classifier-1 is related to training algorithms. It is seen that the recursive LM algorithm is superior to the gradient method in convergence, and has the randomness to reduce the chance of possible trapping due to local minima. In addition, each rule has its own decision (mapping) space, whereas the MFs and the rule weights are directly associated with the characteristics of the decision space. The efficient optimization of classifiers can adjust the boundary characteristics of the decision space so as to reduce misclassifications. This property is especially important for classifier with coarse fuzzy partitions.

The developed integrated classifier generates 3 missed alarms and 7 false alarms, with an overall reliability of 97.6%. Compared with Classifier-2, the integrated classifier can enhance the classification accuracy by properly implementing the future states of the classifier. It follows that adaptively fine-tuning the fuzzy parameters is necessary to enhance the approximation of the mapping from the observed symptoms to the underlying faults. In addition, the fault severity can be recognized because, to some extent, the greater the fault, the more pronounced the feature modulation, and the larger the monitoring indices will become.

The developed integrated diagnostic classifier provides a robust problem solving framework. Machinery conditions vary dramatically in real-world applications, and new system conditions may occur under different circumstances. With the help of an adequate learning algorithm, new information can be extracted from online training, and the diagnostic knowledge base can be expanded automatically to accommodate different machinery conditions.

In general, deterioration history of most machinery components follows a "*U curve*" as illustrated in Fig. 9. It consists of four periods: the run-in stage (I), the normal operation period (II), initial (III) and advanced (IV) failure stages, respectively. Such a trend characteristic is easy for a powerful NF predictor to catch up. If a false alarm is generated during the healthy period II, the false alarm is induced due to noise instead of real defect. Based on the forecast result, the diagnostic state should lie in period III (or initial defect). However random noise will disappear in the following processing steps, and the diagnostic indicator should return to period II (or healthy). Correspondingly, this misclassification can be prevented by the integrated classification /forecasting information. On the other hand, if an object is damaged, its diagnostic indicator should lie in period III (or IV). If a misclassification occurs, or the diagnostic indicator falls in period II, the forecast information will be contradictory to that from the classifier. Comprehensive analysis in Eq. (7) can avoid

this possible missed alarm so as to improve fault diagnostic reliability. In both aforementioned examples, classifier will be updated to accommodate such a noise in the following monitoring applications.
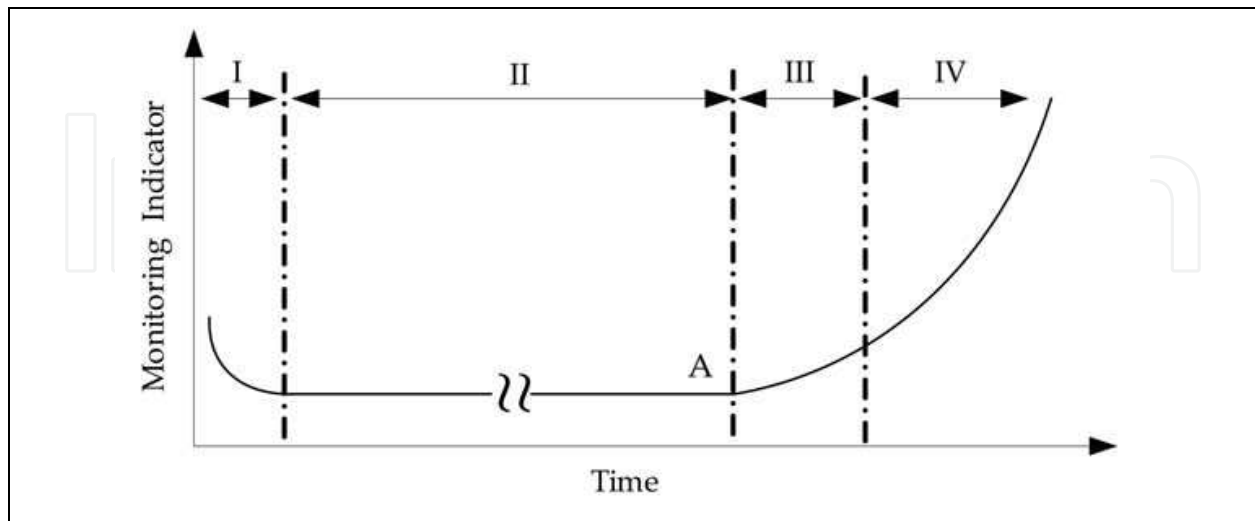

Fig. 9. The deterioration trend of a machinery component.

5. Conclusions

In this paper, an integrated classifier is developed for gear fault diagnostics. The purpose is to provide industries with a more reliable monitoring tool to prevent machinery system performance degradation, malfunction, and sudden failure. The classifier can integrate different features for a more positive assessment of the object's health condition. The diagnostic reliability is improved by properly integrating the future states of the gear, which are forecast by multi-step predictors. An online hybrid training technique based on a recursive LM and LSE is adopted to improve the classifier's convergence and adaptive capability to accommodate different machinery conditions. The viability of the new integrated classifier has been verified by experimental tests corresponding to different gear conditions.

On the other hand, it should be stated that although satisfactory results have been achieved based on the developed integrated classifier, its network architecture is relatively complex which may not be easy for implementation for some real-world applications. Future research is to develop novel evolving fuzzy or neuro-fuzzy classification schemes for more effective diagnostic operations. New training algorithms will be proposed to further improve the training convergence. The proposed techniques will also be employed for real-world industrial applications in vehicles, wind turbines, and manufacturing facilities.

6. Acknowledgement

7. Appendix A: monitoring indices

1. Wavelet energy reference function $R_w(t)$

If $u_r(\tau)$ is the overall residual of the signal average $u(\tau)$, the wavelet energy function is proposed as

$$R_w(t) = \int_{f_1}^{f_2} \left| \int_0^{+\infty} u_r(\tau)\sqrt{s}\, w(t-\tau)\, d\tau \right| ds \qquad (A1)$$

where s and t are the scale (frequency) and time variables, respectively; f_1 and f_2 are the frequency limits of interest. For the gear system in this work, $f_1 = 0.5 f_R N$ and $f_2 = 4.5 f_R N$; $w(t)$ is the mother wavelet, which is a modified Morlet function:

$$w(t) = \exp\left(-\frac{9\ln 2}{\pi^2} f_R^2 t^2 \right) \exp(2\pi k f_R t) \qquad (A2)$$

where $k = 1, 2, \ldots, 5N$.

2. Beta kurtosis reference function $R_b(t)$

The beta kurtosis is the normalized fourth moment of a signal, in terms of the beta function instead of a generally used Gaussian function. If m_t and σ_t^2 represent the mean and variance of a tooth data block, T_d, centered at t, then $R_b(t)$ is defined as the reciprocal of the kurtosis

$$R_b(t) = \frac{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)}{3(\alpha+\beta+1)(2\alpha^2 - 2\alpha\beta + \alpha^2\beta + \alpha\beta^2 + 2\beta^2)} \qquad (A3)$$

where $\alpha = \dfrac{m_t}{\sigma_t^2}(m_t - m_t^2 - \sigma_t^2)$ and $\beta = \dfrac{1-m_t}{\sigma_t^2}(m_t - m_t^2 - \sigma_t^2)$. The derivation of (A3) can be found in [13].

3. Phase modulation reference function $R_p(t)$

For a pair of healthy gears with sound installation and ideal operating conditions, the meshing vibration $u(t)$ can be approximately expressed as,

$$u(t) = \sum_{h=0}^{H} A_h \cos(2\pi h N f_R t + \theta_h) \qquad (A4)$$

where H is the total number of mesh frequency harmonics considered. If a fault occurs in one tooth, because of a change in tooth stiffness, the amplitude and phase functions of the gear meshing vibration will be modulated:

$$u(t) = \sum_{h=0}^{H} (A_h + a_h) \cos[2\pi h N f_R t + \theta_h + \phi(t)] \qquad (A5)$$

The phase modulation $\phi(t)$ can be obtained from the analytical signal of (A5), which is computed by taking the Hilbert transform to $u(t)$. The phase reference function $R_p(t)$ is derived as the maximum phase difference over a tooth period T_d centered at t,

$$R_p(t) = \phi_{\max}(\tau) - \phi_{\min}(\tau), \ \tau \in [\, t - 0.5\, T_d,\ t + 0.5\, T_d\,]. \qquad (A6)$$

8. References

Chelidze, D. & Cusumano, J. (2004). A dynamical systems approach to failure prognosis, *Journal of Vibration and Acoustics*, Vol. 126, 1-7.

Castellano G; Fanelli, A. & Mencar, C. (2004). An empirical risk functional to improve learning in a neuro-fuzzy classifier, *IEEE Transactions on Systems, Man, Cybernetics, Part B*, Vol. 34, 725-731.

Figueiredo, M.; Ballini, R.; Soares, S.; Andrade, M. & Gomide F. (2004). Learning algorithms for a class of neurofuzzy network and applications, *IEEE Transactions on Systems, Man, Cybernetics, Part C*, Vol. 34, 293-301.

Gusumano, J.; Chelidze, D. & Chatterjee, A. (2002). Dynamical systems approach to damage evolution tracking, part 2: Model-based validation and physical interpretation, *Journal of Vibration and Acoustics* , Vol. 124, 258-264.

Ishibuchi, H. & Yamamoto, Y. (2005). Rule weight specification in fuzzy rule-based classification systems, *IEEE Transactions on Fuzzy Systems*, Vol. 13, 428-435.

Isermann, R. (1998). On fuzzy logic applications for automatic control, supervision, and fault diagnosis, *IEEE Transactions on Systems, Man, Cybernetics, Part A*, Vol. 28, 221-235.

Jang, J. (1993). ANFIS: adaptive-network-based fuzzy inference system, *IEEE Transactions on Systems, Man, Cybernetics*, Vol. 23, 665-685.

Korbicz, J.; Koscielny, J.; Kowalczuk, Z. & Cholewa, W. (2004). *Fault Diagnosis: Models, Artificial Intelligence, Applications*, Springer.

Li, L. & Lee, H. (2005). Gear fatigue crack prognosis using embedded model gear dynamic model and fracture mechanics, *Mechanical Systems and Signal Processing*, Vol. 20, 836-846.

Mansoori, E.; Zolghadri, M. & Katebi, S. (2007). A weighting function for improving fuzzy classification systems performance, *Fuzzy Sets and Systems*, Vol. 158, 583-591.

McFadden, P. (1986). Detecting fatigue cracks in gears by amplitude and phase demodulation of the meshing vibration, *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, Vol. 108, 165-170.

Pourahmadi, M. (2001). *Foundation of Time Series Analysis and Prediction Theory*, John & Sons.

Rish, I; Brodie, M.; Ma, S.; Odintsova, N.; Beygelzimer, A.; Grabarnik, G. & Hernandez, K. (2005). Adaptive diagnosis in distributed systems, *IEEE Transactions on Neural Networks*, Vol. 16, 1088-1109.

Tse, P. & Atherton, D. (1999). Prediction of machine deterioration using vibration based fault trends and recurrent neural networks, *Journal of Vibration and Acoustics*, Vol. 121, 355-362.

Uluyol, O.; Kim, K. & Nwadiognu, E. (2006). Synergistic use of soft computing technologies for fault detection in gas turbine engines, *IEEE Transactions on Systems, Man, Cybernetics, Part C*, Vol. 36, 476-484.

Wang, J & Lee, L. (2002). Self-adaptive neuro-fuzzy inference systems for classification applications, *IEEE Transactions on Fuzzy Systems*, Vol. 10, 790-802.

Wang, W. (2008). An intelligent system for machinery condition monitoring, *IEEE Transactions on Fuzzy Systems*, Vol. 16, No. 1, 110-122.

Wang, W.; Ismail, F. & Golnaraghi, F. (2001). Assessment of gear damage monitoring techniques using vibration measurements, *Mechanical Systems and Signal Processing*, Vol. 15, 905-922.

Wang, W. & Vrbanek, J. (2007). A multi-step predictor for dynamic system property forecasting, *Measurement Science and Technology*, Vol. 18, 3673-3681.

Fuzzy Systems

Edited by Ahmad Taher Azar

ISBN 978-953-7619-92-3

Hard cover, 216 pages

Publisher InTech

Published online 01, February, 2010

Published in print edition February, 2010

While several books are available today that address the mathematical and philosophical foundations of fuzzy logic, none, unfortunately, provides the practicing knowledge engineer, system analyst, and project manager with specific, practical information about fuzzy system modeling. Those few books that include applications and case studies concentrate almost exclusively on engineering problems: pendulum balancing, truck backeruppers, cement kilns, antilock braking systems, image pattern recognition, and digital signal processing. Yet the application of fuzzy logic to engineering problems represents only a fraction of its real potential. As a method of encoding and using human knowledge in a form that is very close to the way experts think about difficult, complex problems, fuzzy systems provide the facilities necessary to break through the computational bottlenecks associated with traditional decision support and expert systems. Additionally, fuzzy systems provide a rich and robust method of building systems that include multiple conflicting, cooperating, and collaborating experts (a capability that generally eludes not only symbolic expert system users but analysts who have turned to such related technologies as neural networks and genetic algorithms). Yet the application of fuzzy logic in the areas of decision support, medical systems, database analysis and mining has been largely ignored by both the commercial vendors of decision support products and the knowledge engineers who use them.

INTECH
open science | open minds

www.intechopen.com