A Self-Evolving Fuzzy Classifier for Gear Fault Diagnosis

Ofelia Jianu, Member, IEEE, Wilson Wang, Senior Member, IEEE

Abstract— **Due to the rapid growing market competitiveness, more reliable and robust condition monitoring systems are critically needed in a wide array of industries to improve production quality and reduce cost. As a result, in recent years more efforts have been taken to develop intelligent techniques for online condition monitoring in machinery systems. Although several neural fuzzy classification schemes have been proposed in literature for fault detection, their reasoning architecture remains fixed, allowing only the system parameters to be updated in pattern classification operations. To improve the reliability of machinery fault diagnostics, a self-evolving fuzzy classifier is developed in this work for gear system condition monitoring. The evolution is performed based on the comparison of the potential of the incoming data set and the existing cluster centers. The developed evolving fuzzy classifier has the ability of adding or subtracting rules adaptively. A scaled Kalman filter method is suggested to improve training efficiency. The effectiveness of the developed classifier is evaluated experimentally. Test results show that the proposed classifier can effectively identify the condition of gears, and it outperforms other related methods.**

Index Term— **Evolving classification, gear system condition monitoring, adaptive training.**

I. INTRODUCTION

Gear trains (i.e., gearboxes) are widely used in various types of rotary machines, such as airplanes, ships, automobiles, and manufacturing facilities. Their malfunction can lead to production degradation, malfunction, and even catastrophic failures. The detection of faults in mechanical systems, more precisely in gears, is of great interest nowadays. A reliable online gear monitoring system is critically needed 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.

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: feature (symptom) extraction and pattern classification (diagnosis). Representative features associated with the health condition of a gear train are extracted using the appropriate signal processing techniques. Pattern classification is the process of classifying the features into different categories, each corresponding to a specific machinery health state. Pattern classification can be performed by the use of analytical tools such as numerical models/observers [1] and knowledge-based paradigms [2]. The latter will be utilized in this work because an accurate mathematical model is difficult to derive for a complex gear system, especially when it operates in noisy environment. Knowledge-based diagnostic classification can be performed by reasoning tools such as neural networks [3], fuzzy logic [4], or their synergetic neural fuzzy (NF) schemes [5]. However, most of the current NF classifiers only deal with parameter identification whereas system structure is determined based on expertise and remains unchanged in operations. The alternative is to use cluster-based evolving approach for gradual model development [6]. Several data clustering techniques were suggested, such as K-means clustering and fuzzy C-means clustering. Angelov *et al*. proposed some evolving schemes based on the Takagi-Sugeno fuzzy model, namely eTS, for control and classification applications [7,8]; the cluster structures are determined by a potential measurement and consequent linear parameters are updated by least squares estimate (LSE) algorithm. Despite some promising results, one of the problems in these algorithms is that the spreads and/or centers of the newly generated clusters cannot be updated properly. Song *et al*. suggested a transductive NF inference system with weighted data normalization (namely, TWNFI) for transductive reasoning operations [9]. But compared with eTS tools, TWNFI usually generates more rules in modeling nonlinear systems. In order to tackle these problems and develop a more reliable predictor for real-time industrial applications, the objective of this work is to develop a new self-evolving fuzzy (SEF) technique for fault diagnosis in gear systems. In the proposed classifier, a novel evolving algorithm is proposed to improve the classification efficiency. A novel scaling method is proposed to enhance the covariance matrices in a Kalman

I J E N S

filter to improve training convergence and the stability of the network.

II. THE DEVELOPED SELF-EVOLVING FUZZY CLASSIFIER

The purpose of clustering is to find natural groupings of data in a large data set, thus to revealing pattern in the data so as to provide a concise representation of the data behavior. The SEF system is a self-developing, self-learning fuzzy rule-based system that has both its structures and parameters self-adapting. The SEF structural framework is utilizing the Takagi-Sugeno (TS) fuzzy system with the following form:

$$\Re_j : \text{IF } (x_1 \text{ is } A_1^j) \text{ AND}\ldots\text{AND } (x_n \text{ is } A_n^j) \text{ THEN } y = y_j \quad (1)$$

where $\Re_j$ denotes the jth fuzzy rule (i.e., cluster), $j \in [1, N_r]$, and N_r is the total number of fuzzy rules (clusters); A_i^j is the jth fuzzy set for x_i, $i \in [1, n]$; $y_j = [y_{j1}, y_{j2}, \cdots, y_{jM}]$ is an M-dimensional consequent function [6].

In a fuzzy inference system, nodes in the same layer usually have similar functions. In the proposed SEF classifier, all the fuzzy set membership functions (MFs) are in Gaussian form

$$\mu_{A_i^j} = \exp\left(-\frac{(x_i - m_{ij})^2}{2\sigma_{ij}^2}\right) \quad (2)$$

where m_{ij} and σ_{ij} are the centers and spreads of the MFs, respectively.

If a max-product operator is used for fuzzy reasoning in the premise part, the rule firing strength will be

$$\mu_j = \prod_{i=1}^{n} \mu_{A_i^j}(x_i)$$
$$= \prod_{i=1}^{n} \exp\left(-\frac{(x_i - m_{ij})^2}{2\sigma_{ij}^2}\right) = \exp\left(-\sum_{i=1}^{n} \frac{(x_i - m_{ij})^2}{2\sigma_{ij}^2}\right) \quad (3)$$

After normalization, the overall output indicator will be

$$y = \sum_{j=1}^{N_r} \frac{\mu_j}{\mu_\Sigma} y_j, \qquad j \in [1, N_r] \quad (4)$$

where $\mu_\Sigma = \sum_{j=1}^{N_r} \mu_j = \sum_{j=1}^{N_r} \exp\left(-\sum_{i=1}^{n} \frac{(x_i - m_{ij})^2}{2\sigma_{ij}^2}\right)$

The proposed SEF classifier is a data-driven, non-iterative, one-pass technique modeled off of other potential based algorithms. Different from previous potential based models [7,8], this technique performs a feasibility check on the incoming data sample. It consists of the following procedures.

Step 1: Initialize the initial structure with the first data point $\mathbf{z}_1 = [\mathbf{x}_1 \ \mathbf{y}_1]$: $N_r := 1$; $k := 1$; $P(z_1) := 1$; $\mathbf{m}_1^I := \mathbf{x}_1$; $\sigma_R^I := 0.25$; $\mathbf{m}_R^O := \mathbf{y}_k$; $\sigma_R^O := 0.25$; $\mathbf{m}_1^O := \mathbf{y}_1$

where $\mathbf{m}_1^I$, $\mathbf{m}_1^O$, σ_k^I, σ_k^O are the cluster centers and spreads in the input and output space, respectively, at the k-th time step; $P(z_1)$ is the potential of the first data point.

Step 2: Present next data sample, $z_k = [x_k \ y_k]$, $k := k+1$, where x_k and y_k are the input variables and output variables, respectively.

Step 3: Examine the feasibility of the new data point, which is characterized by the Euclidian distance from the incoming data point to the existing cluster centers:

$$s_k = \left\| z_k - m_i^I \right\| \qquad d_{ij} = \frac{||x_i - \mu_j||}{\Phi_i}$$

where $\Phi_i = \sqrt{\sum_{j=1}^{Q} \frac{1}{Q}(x_i - \mu_j)^2}$, $i = 1, 2, \ldots, N_r$, N_r is the number of clusters generated.

Step 4: Calculate the potential of the new data sample:

$$P_k(z_k) = \frac{k-1}{(k-1)(\theta_k+1) + \gamma_k - 2v_k} \quad (5)$$

where $\theta_k = \sum_{j=1}^{n+1} (z_k^j)^2$; $\gamma_k = \gamma_{k-1} + \sum_{j=1}^{n+1}(\mathbf{z}_{k-1}^j)^2$; $v_k = \sum_{j=1}^{n+1} z_k^j \beta_k^j$;

$\beta_k^j = \beta_{k-1}^j + z_{k-1}^j$

Step 5: Update the potential of the existing clusters:

$$P_k(c_r) = \frac{(k-1)P_{k-1}(c_r)}{k-2 + P_{k-1}(c_r) + P_{k-1}(c_r)\sum_{j=1}^{n+1}(d_{k,k-1}^j)^2} \quad (6)$$

where $d_{k,k-1}^j = z_k^j - z_{k-1}^j$ and $r \in [1, N_r]$.

Step 6: Make the cluster update law based on the potential of the new data sample:

IF $P_k(z_k) > P_k(c_r)$, that is, the potential of new data point is higher than the potential of all the existing cluster centers AND the new data point is close to an existing center, THEN the new cluster center is

$$m_n^I := m_j^I + \frac{\sum(x_k - m_j^I)^2}{N}$$

$$m_n^O := m_j^O + \frac{\sum(y_k - m_j^O)^2}{N}$$

where n denotes a new temporary point, j is an index such that $z_j^* = \arg \min_{i=1}^{N_r} \left\| z_k - m_i \right\|$, $m_n = [m_n^I \ m_n^O]$ and N is the number of samples in the cluster .

Calculate $P_k(m_n)$ or the potential of m_n by Eq. (5).

Update $P_k(c_r)$ the potential of existing clusters by Eq. (6).

IF $P_k(m_n) > P_k(c_r)$, THEN the newly assigned point replaces the cluster center, cluster radius and potential are updated by

$$z_j^* = m_n \; ; \quad \sigma_j^* = \frac{\sum(\sigma_i^* - m_n)^2}{N_r} \; ; \quad P_k(z_j^*) = P_k(z_n)$$

ELSE the data point replaces the cluster center and potential is updated

$$z_j^* = z_k \; ; \quad P_k(z_j^*) = P_k(z_k).$$

Otherwise, IF the potential of new data point is higher than the potential of all existing cluster centers, THEN a new cluster is formed:

$$N_r := N_r + 1 \; ; \; z_{Nr}^* = z_k \, , \; P_k(z_{Nr}^*) = P_k(z_k)$$

Step 7: Update the consequent parameters by recurrent LSE as discussed in the following section.

Step 8: Predict the output at the next time step by Eq. (4)

$$\hat{y}_{k+1} = \psi_k^T \hat{\theta}_k, \quad k = 1, 2, 3, \dots$$

Step 9: Classify the output.

By introducing Step 3 in the identification procedure, the user is able to define the amount of cluster overlapping. In other words, it ensures that a new cluster is not being created if the incoming point is within a cluster. This computation redundancy is being eliminated as a result of introducing the feasibility check.

Different than the method in [7,8], in the proposed SEF technique, a new decision is made before accepting the proposed data set as a cluster center. This step is to ensure that the targeted data set is indeed a better candidate for a cluster center before accepting it. If the targeted data set is not a better candidate for a cluster center, then the target data set is being disregarded and the incoming data set becomes the cluster center.

III. TRAINING OF THE SEF CLASSIFIER

Once the structure is identified using the proposed clustering technique as discussed in Section II, the parameters (both linear and nonlinear) should be properly optimized to provide optimal classification operations. A new scaled Kalman filter (sKF) technique is proposed in this section for system training, especially for nonlinear parameters.

A. The Scaled Kalman Filter Technique

Several training methods have been proposed in the literature for nonlinear parameter optimization, such as gradient algorithm, Levenburg-Marquardt, and Kalman filtering [10]. Based on our systematic investigation, it has been found that the gradient algorithm is prone to being trapped by local minima; whereas Levenburg-Marquardt method cannot be effectively used for large models that may result in oversized variance matrices and significantly slow down the convergence. Among the Kalman filter associated methods, the node decoupled Kalman filter (NDFK) algorithm

outperforms other related Kalman filter related algorithms, which can simplify implementation and reduce memory requirements compared with other Kalman filter related methods. However, the accuracy of the classical NDKF is limited due to its sensitivity to the implementation strategy. The classical NDKF takes two phases in operation, updating and prediction [11]. In the prediction phase, the posteriori state predicted based on the previous time steps is used to produce an estimate of the state at the current time step. In the update phase, the priori prediction is combined with the current information to refine the state estimate as well as the posteriori error covariance matrix. Consider a multivariable system of the form:

$$x_k = F_k x_{k-1} + w_{k-1} \tag{7}$$
$$z_k = H_k x_k + v_k \tag{8}$$

where x_k is a $(n \times 1)$ state vector; F_k is a $(n \times n)$ transition matrix; z_k is a $(r \times 1)$ observation vector; H_k is a $(r \times n)$ observation matrix; w_k is the process noise and v_k is the observation noise, satisfying the following conditions:

$$E(w_k) = E(v_k) = 0$$
$$E(w_k w_i^T) = \begin{cases} Q_k & \text{if } i = k \\ 0 & \text{if } i \neq k \end{cases} \tag{9}$$
$$E(v_k v_i^T) = \begin{cases} R_k & \text{if } i = k \\ 0 & \text{if } i \neq k \end{cases}$$

where $E(\cdot)$ denotes the expectation; Q_k and R_k are the process noise matrix and observation noise covariance matrix, respectively.

In the prediction phase, the predicted state is

$$\hat{x}_{k|k-1} = F_k \hat{x}_{k-1|k-1} \tag{10}$$

Predicted estimate covariance matrix will be

$$P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k \tag{11}$$

where $P_{k-1|k-1} = \text{cov}(x_k - \hat{x}_{k|k})$

In the updated phase, the measurement residual is defined as:

$$\tilde{y}_k = z_k - H_k \hat{x}_{k|k-1} \tag{12}$$

Optimal Kalman gain will be

$$K_k = P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k)^{-1} \tag{13}$$

State estimate update is taken by

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \tilde{y}_k \tag{14}$$

Estimate covariance matrix update is taken by:

$$P_{k|k} = P_{k|k-1} + K_k H_k P_{k|k-1} \qquad (15)$$

where the subscript ($k|k$-1) represents the estimate of the corresponding vector at time instant k given observations up to and including k-1.

The process noise covariance matrix Q and observation error covariance matrix R have considerable impact on the performance of the KF since they are dependent on the application environment and process dynamics [12]. Q and R are responsible for the weight that the system applies between measurements. Consequently, the filter may diverge or never achieve optimum given Q and R exhibit errors. Generally, the covariance matrices are fixed to a value determined from empirical analysis. However, for a complex dynamic system especially in a noisy environment, it is difficult to determine the covariance matrices in advance. Accordingly, a covariance matrix updating technique is proposed in this work to improve the performance of DEKF.

Covariance provides a measure of the strength of the correlation between two or more sets of random variables [13]. In general, when a mathematical model of the system can be obtained, the covariance matrices are chosen based on experience or through experiments. However, this can be a daunting process. It is very difficult (if not impossible) to derive accurate mathematical models for many complex systems in mechanical engineering. This leaves the system to be approximated by a KF. As a consequence, the covariance matrices must also be approximated. This process leaves room for a significant error making the training technique somewhat unreliable. Therefore, a new method to update process noise and observation error covariance matrices is proposed in this section to improve robustness of the training technique, which is defined as:

$$S_k = (H_k P_{k|k-1} H_k^T + R_k)^{-1} \qquad (16)$$

where $P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k$

The update of the process noise and observation error is performed with the help of the innovation covariance matrix:

$$R_{k|k} = R_{k|k-1} + R_{k|k-1}(S_k)^{\beta} \qquad (17)$$

$$Q_{k|k} = Q_{k|k-1} - Q_{k|k-1}(S_k)^{\beta} \qquad (18)$$

where $\beta \in [0\ 1]$ is a design parameter.

The process noise covariance matrix and observation error covariance matrix are diagonal matrices initialized at 0.01 and 0.85, respectively. A series of simulation tests have been performed with initialization values ranging between 0.0001 and 1. After each epoch, the update of the noise covariance matrix and observation error covariance matrix is performed utilizing Eq. (17) and Eq. (18), respectively. Since there is no

direct method to determine the parametric values of these error matrices, the algorithm is modified to account for this task.

During the training with sKF all covariance matrices update to achieve more robust KF. It has been noted that with the introduction of the scaling factor, the predicted estimate covariance matrix, P, changes at a slower rate whereas the process noise covariance matrix and observation error covariance matrix change at a faster rate. Since all covariance matrices are being updated the state estimate update is more robust making the training process more reliable, as it can be noted in the following section.

Many trials have been conducted in order to find the optimal value of β. In this work, a value of $\beta = \dfrac{1}{2N_r}$ will be utilized.

B. Training of the SEF Classifier

The structure identified using the proposed clustering technique consists of both linear and nonlinear parameters which should be optimized properly to provide optimal classification operations. The linear consequent parameters are estimated in the forward pass using the recursive LSE algorithm as stated in [6]. The nonlinear premise parameters are optimized in the backward pass using a new method, scaled Kalman filter (sKF) technique as presented in the this section. The training paradigm employed throughout this work is supervised learning meaning that the input/output data pairs are given and the scope is to find a function that best matches the input/output pair. The goal is to infer how the mapping implied by the data and the cost function is related to the mismatch between the mapping and the data.

IV. PERFORMANCE EVALUATION

The developed SEF classifier is implemented in this section for gear condition monitoring. This work will focus on localized gear fault.

A. Experimental Apparatus

The experimental setup used in this work is shown in Fig. 1. It is driven by a 3HP DC motor with a speed range from 20 to 4200 RPM. The rotational speed of the tested system is controlled by a speed controller. Vibration signals are collected using ICP accelerometers with a sensitivity of 100mV/g. The accelerometers are secured on the gearbox casing along different orientation. The reference signals are generated using a high-speed retro optical sensor. The collected signals are fed to the computer through a data acquisition board which has built-in A/D converters and antialiasing filters. Load is applied by a magnetic brake system.
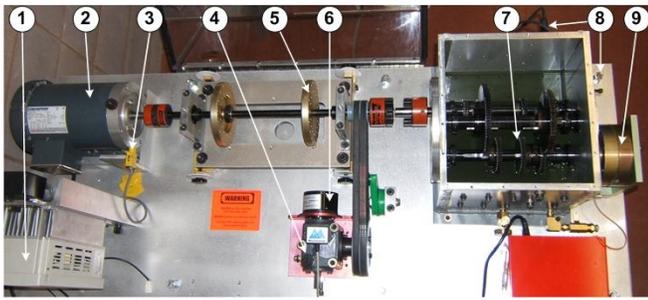
Fig. 1. Machinery dynamics simulator: 1-speed controller, 2-drive motor, 3-optical sensor, 4-bevel gearbox, 5-imbalance disks, 6-small load drive, 7-tested gears, 8-accelerometers, 9-load drive.

B. Monitoring Indexes

In this work, the condition monitoring of the gear system is conducted gear by gear. The measured vibration from the experimental setup is an overall signal associated with different vibratory sources, such as magnetic break, shafts, bearings, and gear mesh. Each gear signal will be differentiated by applying a time synchronous average filtering process [14]. As a result, the signals nonsynchronous to the rotation of the gear of interest can be removed, and each gear signal is derived and represented in one full revolution (i.e., the signal average).

In condition monitoring, the monitoring indices should be sensitive to pattern modulation due to machinery faults but insensitive to noise. Several signal processing techniques have been proposed in the literature for gear fault diagnosis. Based on processing tools, these techniques can be classified into methods in the time-domain, frequency domain, and time-frequency domain, respectively. But each has its own advantages and disadvantages. In this case, three monitoring indexes are selected for gear fault diagnosis: energy operator, kurtosis, and crest factor.

a) Energy operator is related to the energy of the signal. If a gear is damaged, usually its vibration signal will be modulated, and the signal energy will be increased. Consider a gear signal s, which is the signal average obtained by the time synchronous average filtering, the energy operator E_O is defined as:

$$E_O = \frac{N_O^2 \sum_{i=1}^{No} (\Delta s_i - \Delta \bar{s})^4}{\left(\sum_{i=1}^{No} (\Delta s_i - \Delta \bar{s})^4 \right)^2} \tag{19}$$

where $\Delta \bar{s}$ is the mean value of signal Δs; N_O is the number of points in dataset s and $\Delta s = s_{i+1} - s_i$ where s_i is the ith data point of the signal average of the gear of interest.

b) Kurtosis is the forth moment of the normalized signal. Whenever a gear is damaged, some high frequency peaks are generated in the signal. Kurtosis is a measure of the peakness in the signal, which will modulate the tail property of the signal's probability distribution function. The kurtosis K_U is defined as:

$$K_U = \frac{N_O^2 \sum_{i=1}^{No} (s_i - \bar{s})^4}{\left(\sum_{i=1}^{No} (s_i - \bar{s})^2 \right)^2} \tag{20}$$

where $\bar{s}$ is the mean value of signal s; the denominator is the fourth moment of the signal standard deviation, which is for normalization.

c) Crest factor is a measurement of peak-to-peak value of the signal. If a defect occurs and/or the defect size varies, the peak values will change correspondingly, which is a good health indicator for some severe gear fault. The crest factor C_F is defined as:

$$C_F = \frac{s_{\max} - s_{\min}}{\sqrt{\frac{1}{N_O} \sum_{i=1}^{No} s_i^2}} \tag{21}$$

where the denominator represents the root mean square error of the signal, which is used for normalization; $s_{\max}$ and $s_{\min}$ are the maximum and minimum values of the gear signal, respectively. Details related to the signal processing techniques and post processing strategies can be found from [15.16].

C. Gear Fault Diagnosis using the Developed SEF Classifier

Various faulted gears have been tested, with both spur gears and helical gears, corresponding to the common defect cases such as severe cracked gears, minor cracked gears, and of course, healthy gears. Although many tests have been performed under different shaft speeds and loading conditions, one example will be discussed in this paper to demonstrate how to implement the developed SEF classifier for automatic gear fault detection. Fig. 2 shows an example of the tests with a spur gear with a 20% missing tooth. The illustration will be based on the operating conditions of shaft speed of 15Hz with the loading level of 150mA.
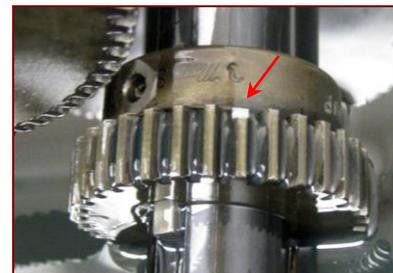

Fig. 2. The broken tooth on a spur gear.

The proposed technique is a combination of the identification technique SEF and training technique sKF, denoted as SEF-sKF. To test the effectiveness of the proposed sKF training technique, the test results of the method called SEF-NDKF is listed, which is the paradigm employing the proposed clustering technique but using a general NDKF. Furthermore, two of well-accepted clustering methods, eTS [7] and TWNFI [9] trained by NDKF, are also used for comparison.

In classification analysis the same condition indices computed above are used as inputs to the related classifiers. The network has one output, which is further classified into two categories as follows:

$\Re_1$: (IF Output<0.5) THEN (Gear is Faulty with weight w_1)

$\Re_2$: (IF Output $\geq$ 0.5) THEN (Gear is Healthy with weight w_2)

The classifiers are trained over 50 epochs using the gear data sets with η_r = 0.85 and η_q = 0.001. A total of 85 input/output data pairs corresponding to different load/speed levels are computed using the related techniques. Fifty data pairs are used for cluster generation. Test results of the related classifiers are summarized in Table I, where Fig. 3 illustrates convergence of the related classifiers in terms of the root mean square error (RMSE). It is clear that the proposed SEF-sKF technique outperforms other related classifiers. It generates only two clusters (or MFs) and with the best performance in terms of accuracy (with the least RMSE) and efficiency (with the shortest operation time per epoch).

In comparing the proposed SEF scheme with the classical methods used in sTS and TWNFI, the SEF has a higher classification accuracy. In comparison of the SEF-sKF and the SEF-DEKF, the proposed sKF method can improve training efficiency and convergence.

The clustering processes using the proposed SEF-sKF technique are demonstrated in Figures 4 to 6, whereas the identified classifier network is illustrated in Fig. 7.

Fig. 4. Identified clusters and membership functions (KU vs. EO)

Table I
Spur gear test results

Classifier Scheme	Number of Clusters	Success Rate (%)	Average RMSE	Average Time per Epoch (s)
TWNFI	5	93.7	0.328	0.082
eTS	4	95.8	0.315	0.067
SEF-NDKF	2	95.4	0.316	0.042
SEF-sKF	2	96.1	0.313	0.042

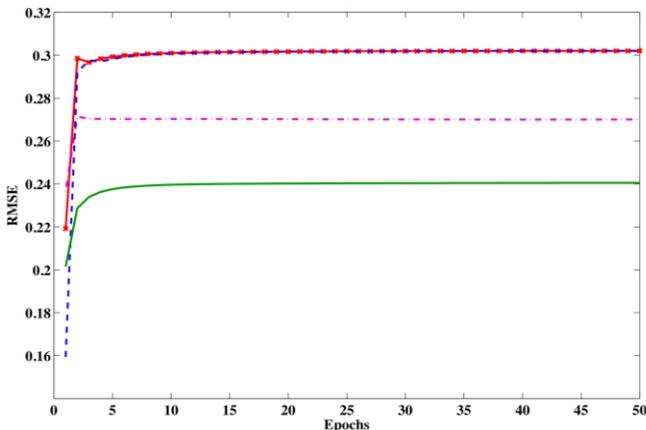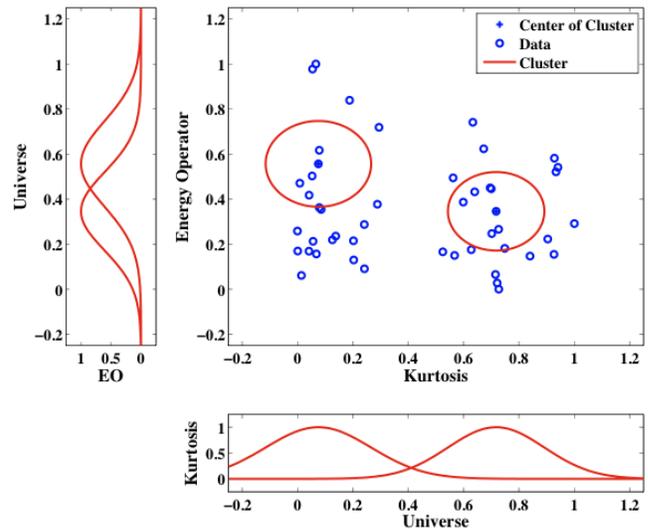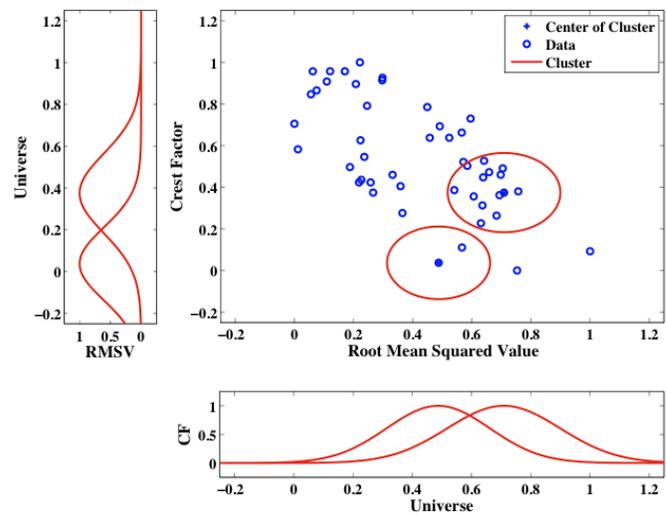Fig. 5. Identified clusters and membership functions (RMSV vs. CF)

Fig. 3. Root mean square error comparison: Green line: SEF-sKF; Purple dotted line: SEF-NDKF; Red-dotted line: eTS; Blue-dotted line: TWNFI
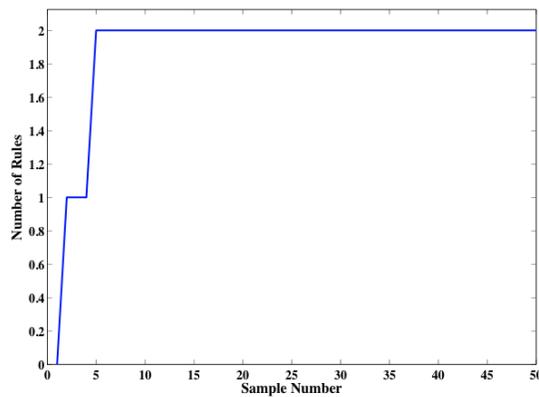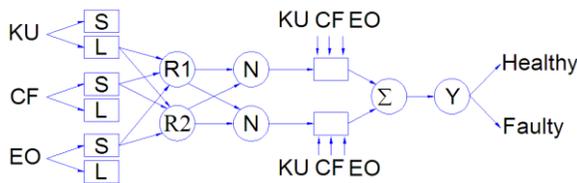
Fig. 6. Evolution of the clusters

Fig. 7. Identified classifier network structure by using SEF-sKF.

V. CONCLUSIONS

A self-evolving fuzzy, SEF, classifier has been developed in this paper for gear fault diagnostics. The evolution is performed based on the comparison of the potential of the incoming data set and the existing cluster centers. The suggested self-evolving algorithm has the ability of adding or subtracting rules adaptively and modifying existing rules and parameters. The distinguishable patterns can be determined between the input data and the output patterns. A novel scaled Kalman filter, sKF, technique is suggested to enhance covariance matrices during the training process to improve training efficiency and convergence. The effectiveness of the proposed classifier has been verified experimentally. Test results have shown that the proposed SEF-sKF classifier can effectively diagnose the health conditions of gear systems, and it outperforms other related methods.

REFERENCES

[1]   K. Aslantas and S. Tasgetiren, "A study of spur gears pitting formation and life prediction," *Wear*, Vol. 257, pp. 1167-1175, 2004.

[2]   W. Wang and O. Jianu, "A smart monitor for machinery fault detection," *IEEE/ASME Transactions on Mechatronics*, Vol. 15, No. 1, pp. 70-78, 2010.

[3]   J. Jang, C. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing*, Prentice Hall, 1997.

[4]   G. Vachtsevanos and F. Lewis, *Intelligent Fault Diagnosis and Fault Prognosis for Engineering Systems*, John Wiley & Sons, 2006.

[5]   W. Wang, "An enhanced diagnostic system for gear system monitoring," *IEEE Transactions on Systems, Man, Cybernetics, Part B*, Vol. 34, pp. 102-112, 2008.

[6]   W. Wang, D. Li and J. Vrbanek, "An adaptive evolving technique for system dynamic state analysis," *Journal of Neurocomputing*, Vol. 85, pp. 111-119, 2012.

[7]   P. Angelov, D. Filev, "An approach to online identification of Takagi-Sugeno fuzzy models," *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics*, Vol. 34, No. 1, pp. 484-498, 2004.

[8]   P. Angelov, X. Zhou, "Evolving fuzzy-rule-based classifiers from data streams," *IEEE Transactions on Fuzzy Systems*, Vol. 16, No. 6, pp. 1462-1474, 2008.

[9]   Q. Song and N. Kasabov, "TWNFI - A transductive neuro-fuzzy inference system with weighted data normalization for personalized modeling," *Neural Networks*, Vol. 19, pp.1591-1596, 2006.

[10] Y. Yang and W. Gao, "Comparison of adaptive factors in Kalman filters on navigation results," *Journal of Navigation*, Vol. 58, pp. 471-478, 2005.

[11] H. Li, D. Wunsch, E. O'Hair, M. Giesselmann, "Extended Kalman filter training of neural networks on a SIMD parallel machine," *Journal of Parallel and Distributed Computing*. Vol. 62, No. 4, pp. 544-562, 2002.

[12] W. Ding, J. Wang, C. Rizos, "Improving covariance based adaptive estimation for GPS/INS integration," *Journal of Navigation*, Vol. 60, No. 3, pp.517-529, 2007.

[13] Q. Jackson, D. Landgrebe, "An adaptive method for combined covariance estimation and classification," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 40, No. 5, pp. 1082-1087, 2002.

[14] W. Wang, F. Ismail, F. Golnaraghi, " Assesment of gear damage monitoring techniques using vibration measurements," *Mechanical Systems and Signal Processing*, Vol. 15, pp. 905-922, 2001.

[15] L. Combet, L. Gelman, "An automated methodology for permorming time synchronous averaging of a gearbox signal without speed sensor," *Mechanical Systems and Signal Processing*, Vol. 21, No. 6, pp. 2590-2606, 2007.

[16] J. Dron, F. Bolaers, L. Rasolofondraibe, "Improvement of the sensitivity of the scalar indicators (crest factor, kurtosis) using a de-noising method by spectral subtraction: application to the detection of defects in ball bearings," *Journal of Sound and Vibration*, Vol. 270, No. 1-2, pp. 61-73, 2004.