Expert Systems with Applications 37 (2010) 8831–8840

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Intelligent condition monitoring and prognostics system based on data-fusion strategy

Gang Niu [a], Bo-Suk Yang [b,*]

[a] China Aero Polytechnology Establishment, Chaoyang-district, Beijing, 100028, PR China
[b] School of Mechanical Engineering, Pukyong National University, San 100, Yongdang-dong, Nam-gu, Busan 608-739, South Korea

ARTICLE INFO

Keywords:
Data fusion
Condition monitoring
Alarm setting
Data-driven prognostics
Degradation assessment
Remaining useful life prediction

ABSTRACT

This paper proposes an intelligent condition monitoring and prognostics system in condition-based maintenance architecture based on data-fusion strategy. Firstly, vibration signals are collected and trend features are extracted. Then features are normalized and sent into neural network for feature-level fusion. Next, data de-noising is conducted containing smoothing and wavelet decomposition to reduce the fluctuation and pick out trend information. The processed information is used for autonomic health degradation monitoring and data-driven prognostics. When the degradation curve crosses through the specified threshold of alarm, prognostics module is triggered and time-series prediction is performed using multi-nonlinear regression models. Furthermore, the predicted point estimate and interval estimate are fused, respectively. Finally, remaining useful life of operating machine, with its uncertainty interval, are assessed. The proposed system is evaluated by an experiment of health degradation monitoring and prognostics for a methane compressor. The experiment results show that the enhanced maintenance performances can be obtained, which make it suitable for advanced industry maintenance.

1. Introduction

With the development of industry, especially heavy industry in nuclear power, automobile, shipbuilding and aircraft, which take on important roles in modern society, condition-based maintenance (CBM) technology shows increasing importance due to its effective roles in improving plant production availability, reducing downtime cost and enhancing operating reliability. The open system architecture for condition-based maintenance organization (OSA-CBM) (Hall & Llinas, 1997; Thurston & Lebold, 2001) has divided a CBM system into seven different layers, with technical modules solution as shown in Fig. 1. The core functions among the architecture can be summarized as condition monitoring, health assessment and prognostics ranging from layer 3 to layer 5.

Condition monitoring involves comparing on-line data with expected values; if necessary it should be able to generate alerts based on preset operational limits. Health assessment serves prescribing if the health of the monitored component or system has degraded, and exerting fault diagnosis. The primary tasks of the prognostics module deal with calculating the future health of an asset and report the remaining useful life (RUL) (Thurston & Lebold, 2001). In reality, however carrying a reliable and effective CBM technology is a challenge due to not only the outer electromagnetism-noise but the inner complex structure of the machine, other than the abstruse failure mechanisms.

Currently, another two developing research areas are intelligence and data-fusion technology. Intelligence system means that any formal or informal system to manage data gathering, to obtain and process the data, to interpret the data, and to provide reasoned judgments to decision makers as a basis for action (Bengtsson, 2004). Here, intelligent monitoring indicates that using artificial intelligence (AI) techniques to continuously monitor or automatically detect health condition of the machine. While intelligent prognostics is defined as a systematic approach that can continuously track health degradation and extrapolating temporal behavior of health indicators to predict risks of unacceptable behavior over time as well as pinpointing exactly which components of a machine are likely to fail (Lee, Ni, Djurdjanovic, Qiu, & Liao, 2006).

Data-fusion techniques, according to fusion contents, basically include signal-level fusion, feature-level fusion and decision-level fusion (Hall & Llinas, 1997). Applying fusion techniques into engineering practice has been receiving increasing attentions in recent years. Especially, with the rapid progress of advanced sensor and signal processing technologies, fusing large of mutual information becomes possible, which is expected to bring about enhanced CBM performances. A number of fusion techniques have been identified in improving accuracy of machinery faults diagnosis, for example, engine fault diagnosis using Dempster–Shafer evidence theory (Basir & Yuan, 2007), motor fault diagnosis using multi-agent fusion (Niu, Han, Yang, & Tan, 2007), tank reactor diagnosis using multiple

* Corresponding author. Tel.: +82 51 629 6152; fax: +82 51 629 6150.
E-mail address: bsyang@pknu.ac.kr (B.-S. Yang).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.06.014

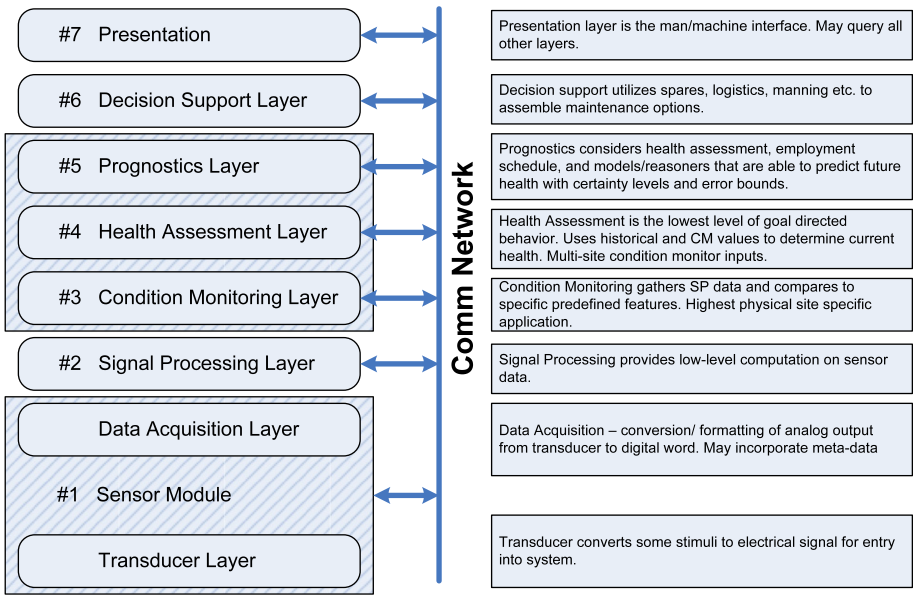


Fig. 1. Open system architecture for condition-based maintenance (OSA-CBM).

neural networks fusion (Zhang, 2006) and cutting tool diagnosis using fuzzy fusion (Kohonen, 1995). However, the applications of data-fusion technology in machinery condition monitoring and prognostics have not received sufficient attentions yet, and relevant cases are seldom.

As far as real-time condition monitoring is concerned, indicating potential failures relies on reliable degradation indicator and appropriate alarm setting. Failures can often be attributed to many correlated degradation processes, which could be reflected by multiple degradation indicators extracted from sensor signals (Lee et al., 2006). Each degradation indicator has its own merits and shortcomings and is only effective for certain failure at certain stage. Therefore, an effective fusion of multiple indicators may potentially provide an improved approach to degradation monitoring. Also, alarm setting determines a division between normal and degradation operating condition of machinery, which is important so that a slight change about a threshold can mean a drastic change in false alarms, missed detections, and prognostics prediction time horizons. The threshold value usually can be set up according to the international standards (ISO 13381-1, ISO 10816 and ISO 7919) with historic experiences of monitored machine for proper adjustment. However, the ISO standards are only limited in root mean square (RMS) indicator of vibration signals, and just for general size and mounting of the machinery such as pump, electric motor, compressor and turbine, which limit prevalent application of the international standards. Even if different indicators are used, each one has its own threshold setting scheme and making alarm setting difficult. Therefore, a fused degradation indicator also needs a new corresponding alarm setting mechanism.

As far as prognostics is concerned, the accurate assessment of RUL is difficult to achieve due to a number of reasons such as suitable data for training, prediction models, and uncertainty management. Fusion information for prognostic purposes is a fairly new endeavor and will likely lead to the development of new techniques that are specialized to perform related tasks (Goebel, Bonanni, & Eklund, 2005).

In this paper, a novel intelligent condition monitoring and prognostics system is introduced based on data-fusion strategy for enhancing CBM quality which contains two parts: on-line condition monitoring module and data-driven prognostics module. In on-line condition monitoring module, raw data is initially collected from multiple sensors and trend features of signals are extracted. The obtained feature values are then normalized and grouped as input of a feature-level fusion. Next, a de-noising step is conducted for the fused result, and the processed value is taken as degradation monitoring indicator. Finally, alarm value is set up, and condition monitoring is carried out. When the degradation indicator reaches at the threshold of alarm setting, a data-driven prognostics module is triggered. At first, the time-series monitoring data sets are reconstructed, then nonlinear regression models are employed to predict future degradation trajectory of machine health state. Furthermore, the predicted values and bias from different models are fused to enhance the reliability. Finally, remain useful life and its uncertainty interval can be assessed.

The rest parts of this paper are organized as follows. In Section 2, the proposed system is introduced in detail; the basic concept of each part is explained. Section 3 briefly introduces some background knowledge used in this paper. Section 4 describes a practical case of condition monitoring and prognostics for a methane compressor based on data-fusion strategy to demonstrate the effects of the system. At last, conclusions are clarified in Section 5 on the basis of the experiment results in Section 4.

2. A proposed condition monitoring and prognostics system based on data fusion

The technology of CBM has made a number of progresses in recent years; however, many fundamental issues still remain (Jardine, Lin, & Banjevic, 2006; Lee et al., 2004):

- Indicators, used for accurate condition monitoring and prognostics, need to be developed.
- Efficient and fast signal processing algorithms need to be developed.
- Alarm setting based on ISO standards is only available to RMS feature, new alarm setting techniques need to be developed for other degradation indicators.
- Currently, methods are generally focused on solving the failure prediction problems; tools for system performance assessment and degradation prediction have not been well addressed.
- Fast and precise prognostic approaches need to be developed.

In this paper, a data-fusion based condition monitoring and prognostics system is proposed for solving the above issues. The flowchart of the proposed system is shown in Fig. 2. The proposed system is based on OSA-CBM architecture and focused on the condition monitoring module and prognostics module. Data-fusion strategy is emphasized and employed in the whole system for enhancing effect and quality of machine health assessment. The procedures of the proposed system can be summarized as follows.

2.1. On-line condition monitoring module

Condition monitoring determines where a system or component is on the indication curve. Is it "nominal"? Does some "anomaly" condition exist? Or, is it somewhere between those two extremes? Determining where we are on the health curve is the first step in prognostics.

The advanced equipment and sensor technologies have provided more and more immediate data to reveal a machine's condition. However, such data have not been effectively analyzed and put to use in practice due to their enormous volume and the lack of efficient analysis methods. If we can establish a mechanism to analyze these real-time data, the equipment condition can be then observed and evaluated in a more timely fashion (Chen & Wu, 2007).

In this module, at first, signals of multi-sensors attached on operating machine are collected and features embossing operating state are extracted. Then those calculated features are normalized and grouped as input set for a feature-level fusion algorithm. The enormous mutual information reflecting machine health is expected to generate a robust health degradation indicator for machine monitoring, diagnosis and prognostics. Self-organizing map (SOM) neural network is chosen as a feature-level fusion algorithm. The fused output, minimum quantization error (MQE), has the merits of unsupervised learning, short training time, easy operation and consistent tracking performance.

Next, the process of de-noising is considered for filtering process noise of features extraction and fusion. The methods of smoothing and wavelet decomposition are suggested. Smoothing is an easy but efficient way, which can capture important patterns in the data, while leaving out noise. Furthermore, wavelet decomposition is performed to delete detail noise and increasing smoothness. After the de-noised process, a clear tracking trend of operating state is picked out. Next, an automatic alarm setting strategy is suggested based on the longest time constant of machine and statistical properties of the candidate baseline, which can be employed to solve the problem of non-RMS indicator alarm setting. Finally, condition monitoring can be carried out and a comparison is exerted continuously between the alarm threshold and each testing MQE value. If the monitored indicator reaches to the threshold, a data-driven prognostics module would be triggered.

2.2. Data-driven prognostics module

According to the definition in ISO 13381-1, prognostics is an estimation of time to failure and risk for one or more existing and future failure modes, and is normally intuitive and based on experience. Approaches for prognostics reasoning can be classified into four categories (Dong & He, 2007):

- Physical models;
- Rule-based or case-based systems;
- Model-driven statistical learning methods;
- Data-driven statistical learning models.

Considering the universality and convenience, data-driven models are employed in this research compared with other approaches which mainly rely on accurate mathematic or physic modeling to special equipment.

Usually performance degradation trend is reflected as a nonlinear or chaotic character. Therefore, state space reconstruction becomes the first step in nonlinear chaotic time-series prediction. In this system, the reconstruction parameters, delay time and embedding dimension are selected by C-C method and false nearest neighbor (FNN) method, respectively.

Then degradation prediction is exerted. In this process, two models, Dempster–Shafer regression (DSR) and least square-support vector machine (LS-SVM) are especially appropriate for nonlinear time-series prediction thus proposed for data-driven prognostics. In addition, the two models can give predictions of not only degradation trend (or point estimate) but also uncer-

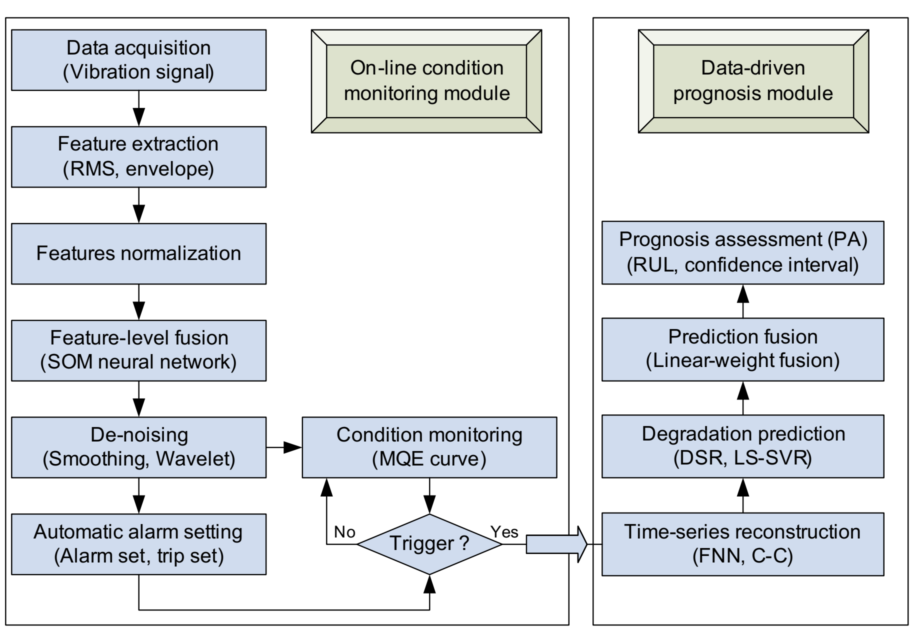


Fig. 2. The flowchart of proposed system.

tainty bounds (or interval estimate). Next, combining the prediction results of DSR and LS-SVM is conducted by a linear-weight fusion method. At this step, other advanced fusion methods are not proper because the final fused results should be accordance with alarm setting and trip setting on unit and scale. The weights are tuned by experience. Finally, prognostics assessment (PA) is carried out. The cores of PA is estimating the remaining useful lifetime of a failing component or system and assigning uncertainty bounds to the degradation trend that will provide maintainers with the earliest and the latest (with increasing risk) time to perform maintenance and the associated risk factor when maintenance action is delayed.

3. Description of relevant background knowledge

This section covers a brief introduction of the relevant knowledge used in the proposed system. The materials contain data-fusion techniques, de-noising methods, automatic alarm setting strategy and time-series prediction models.

3.1. Data-fusion

A "fused" definition, which fits many examples in engineering, is identified as the process of combining data and knowledge from different sources with the aim of maximizing the useful information content, for improved reliability or discriminant capability, whilst minimizing the quantity of data ultimately retained (http://www.data-fusion.org/). A common objective of data fusion from an equipment health management perspective is to combine relevant system information in the most efficient manner possible for improving the confidence and accuracy of diagnostic and prognostic approaches (Dong & He, 2007). In this paper, two standard data-fusion methods are employed.

3.1.1. Degradation indicator using SOM neural network fusion

The SOM is a neural network concept developed by Boutros and Liang (2007). It forms a one or two-dimensional presentation from multi-dimensional data. The topology of the data is kept in the presentation such that data vectors are located next to each other on the map. Usually, the input features are normalized firstly and the SOM is trained iteratively. In each training step, one sample vector **X** from the input feature set is chosen randomly and the distance between it and all the weight vectors of the SOM is calculated using some distance measure such as Euclidian distance. Next the best matching unit (BMU) can be identified whose weight vector is closest to **X**. Then the weight vectors of the BMU as well as its topological neighbors are updated so that they are moved closer to the input vector in the input space. At the end of the learning process, the weight vectors are grouped in clusters depending on their distance in the input space.

The condition of the machine can be described by its matching region in SOM. And the operation state changes can be described by the trajectory of its BMUs in SOM. In normal operation, the BMUs should follow well-defined paths or trajectories in normal regions. When an incipient fault appears, its BMUs would deviate from the normal region. By plotting the trajectory of current data on a labeled map, the machine condition can be followed over time. If a probability of the next BMUs is available, a prediction of the next possible machine state can be assessed. Furthermore, Qiu, Lee, Lin, and Yu (2003) put forward that degradation monitoring can be based on the calculation of minimum quantization error (MQE) of the new measurement data to an SOM trained using normal operation data sets.

The SOM is first trained with normal operation data. Then the feature vector corresponding to the unidentified measurement is compared with the weight vectors of all map units. If the smallest difference exceeds a predetermined threshold, the process is probably in a fault situation. The distance between the BMU and the input data actually indicates how far the input data deviate from the region of normal operation. The condition monitoring considers MQE as fused degradation indicator, which can be defined as

$$MQE = \|D - m_{BMU}\| \tag{1}$$

where D is the input data vector and m_{BMU} stands for the weight vector of the BMU. Therefore, the condition degradation can be quantized and visualized by following the trends of MQE.

3.1.2. Time series prediction using linear-weight fusion

Prognostic is a key element in equipment health management by providing an estimate for remaining equipment life. Accurate prognostics is difficult to achieve for a number of reasons. The range from the availability of suitable data sets for training of algorithms to the unsolved issue of validating prognostic approaches to the incorporation of future usage information to the need to manage uncertainty to the need to find technology that aggregates diverse information sources and returns a continuous output reflective of remaining life (Goebel & Bonissone, 2005).

For data-driven prognostics, using one type of prediction model is hard to get satisfied prediction result, thus multi-prediction fusion affords a potential possibility in improving the prediction effect. Fusion in this process is already not suitable for using complex methods, considering the final fused results should be accordance with alarm setting and trip setting on unit and scale. Hence, the simple linear-weight fusion is employed.

The weighted linear fusion (combination or average) uses the factors standardized to a continuous scale of suitability from 0 (the least suitable) to 1 (the most suitable). The weights can be determined by experience or optimization techniques such as genetic algorithm (Roemer & Kacprzynski, 2001).

3.2. Indicator de-noising

Though raw signals are usually collected with some simple signal preprocessing, new noises may be produced in followed processes of feature extraction and fusion. Therefore, indicator de-noising is needed, or else, the degradation curve is often fluctuant so that easily triggers a false alarm and leads to a cost-monitoring.

3.2.1. Smoothing: moving average

Smoothing a data set is to create a function that attempts to capture important patterns in the data, while leaving out noise. There are two types of smoothing algorithms: moving average and local regression. The most common one is the *moving average* often used to try to capture important trends in repeated statistical surveys; therefore, this method is considered and operated by averaging a number of points from the input signal to produce each point in the output signal. In equation form, this is written:

$$y[i] = \frac{1}{M}\sum_{j=0}^{M-1} x[i+j] \tag{2}$$

where $x[i]$ is the input signal, $y[i]$ is the output signal and M is the number of points in the average.

3.2.2. Wavelet de-noising

Wavelet de-noising method is based on the principle of multiresolution analysis. By using multi-level wavelet decomposition, the discrete detail coefficient and approximation coefficient can be easily obtained. Small detail coefficients are assumed to be dominated by noise and carry little information. Replacing these coefficients by zero eliminates a major part of the noise without affecting the signal very much (Shao & Nezu, 2005). Donoho (1995) proposes the following scheme for de-noising:

(a) *Decomposition*: Choose a wavelet and a level N. Compute the wavelet decomposition of the original signal at level $N \cdot Y = Wy$.
(b) Detail coefficients thresholding in the wavelet domain, according to so-called hard thresholding

$$\widehat{X} = T_h(Y,t) = \begin{cases} Y, & |Y| \geqslant t \\ 0, & |Y| < t \end{cases} \tag{3}$$

or according to so-called soft thresholding

$$\widehat{X} = T_s(Y,t) = \begin{cases} \mathrm{sgn}(Y)(|Y|-t), & |Y| \geqslant t \\ 0, & |Y| < t \end{cases} \tag{4}$$

(c) *Reconstruction*: Compute wavelet reconstruction based on the original approximation coefficients of level N and the modified detail coefficients of levels from 1 to N.

$$\hat{x} = W^{-1}\widehat{X}$$

where y is the noisy observations, t is threshold. Capital letters denote variables in the transform domain. Let W be a left invertible wavelet transformation matrix of the discrete wavelet transform (DWT). The estimate $\widehat{X}$ is obtained by simply keeping or zeroing the individual wavelet coefficients and the result $\hat{x}$ is the recovered signal.

3.3. Automatic setting of alarms

A formal approach can be summarized in engineering norms such as ISO 10816, which suggest acceptable levels of vibration according to machinery size and mounting. Furthermore, the practical alarm value can be tuned according to the user experience. However, one of the outstand drawbacks in the ISO 10816 is considering RMS feature as the only monitoring indicator, which limits the application of data-fusion technique in improving accuracy of condition monitoring.

Recently, Ginart, Barlas, Goldin, and Dorrity (2006) put forward a new alarm setting mechanism based on the longest time constant of machine and statistical properties of the candidate baseline, which can be employed to solve the problem of fused indicator alarm setting. Firstly, baseline candidate is decided by calculating mean and deviation of fused indicator data set in good operational and health condition. Then the indicator is checked and an alarm coefficient is determined according to Table 1, which is established according to extensive references from heuristics, norms, basic mathematical structure of acceleration model, ration SD/mean and slope of baseline etc. Finally, the alarm setting is finished by multiplying an alarm coefficient with the established baseline. Two criteria are considered in checking the acceptance of monitoring indicator, and determining alarm coefficient as follows:

3.3.1. Criteria based on largest time constant

The largest time constant of the system is related with the size of the machine and the processes themselves. For large industrial equipment such as huge pumps, blowers and centrifuges, 20 min is usually a very conservative estimation. In general terms, it is reasonable to establish a time window t that is at least three times of the largest time constant of the system. This allows the system hearing up fully and reaching at least a quasi-steady state, which can allow only changes coming from the failure not from the process itself.

The response for a step functions in the first order system is:

$$c(t) = 1 - e^{-t/\tau} \tag{5}$$

where τ is the time that take for any first order system to reach 67% of the final steady state (Ogata, 1996). Its slope of Eq. (5) can be computed as:

$$\frac{\partial c(t)}{\partial t} = \frac{1}{\tau} e^{-t/\tau} = m \tag{6}$$

The first criterion for selecting a baseline in mathematical terms is:

$$|m| < 0.017 \tag{7}$$

where m is the slope of the linear regression of the candidate baseline.

3.3.2. Criteria based on statistics (6σ) (Nelson, 2004; O'Connor, 2002)

This criterion is based on the very low probability, assuming Gaussian distribution, of obtaining a consistence value greater than six times the standard deviation (6σ) normalized by the mean (μ) of the indicator value. This restriction minimizes the possibility down to very low random probabilities of having a false alarm when the machine is in good operational and health condition (Byington, Watson, Kalgren, & Safa-Bakhsh, 2003). The acceptable baseline usually complies with the following relation:

$$k = \frac{\sigma}{\mu} < \frac{1}{6} \tag{8}$$

After getting acceptable parameters m and k, the alarm coefficient can be indexed from Table 1. Then proper alarm setting can be calculated as:

$$Alarm_{mean(deviation)} = Alarm\ coefficient \times Baseline_{mean(deviation)} \tag{9}$$

Table 1
Proposed alarm coefficients.

	$\alpha/\mu < 0.04$	$\alpha/\mu < 0.082$	$\alpha/\mu < 0.167$
$m < 0.004$	2	3	4
$m < 0.0082$	3	4	5
$m < 0.0167$	4	5	5

3.4. Nonlinear prediction models

3.4.1. Dempster–Shafer regression (DSR)

DSR or evidence regression (EVREG) was introduced by Petit-Renaud and Denoeux (2004) who adopt the subjectivist and non-probabilistic view of Smets' transferable belief model (TBM) (Smets, 1998; Smets & Kennes, 1994). Basically, the method considers each training sample in the neighborhood of the input vector $\mathbf{x}$ as a piece of evidence regarding the value of the output y. The pieces of evidence are discounted as a function of their distance to $\mathbf{x}$, and pooled using Dempster's rule of combination.

This model is described as follows:

(1) Calculate mass function m (or basic probability assignment) in training set ζ.
Given $\mathbf{x}$ as an input vector, and y as the corresponding unknown output. The evidence samples $e_i = (\mathbf{x}_i, m_i)$ in the neighborhood of the input vector $\mathbf{x}$ are sources of relevant information on the response output y.
(2) Determine fuzzy belief assignment (FBA) on y denoted by $m_y[\mathbf{x}, e_i]$.
Considering the potential imprecision and uncertainty in learning set information, DSR takes the form of a FBA. Each element $e_i = (\mathbf{x}_i, m_i)$ of the training set is a piece of evidence concerning the possible value of y_i, which can be represented by a FBA $m_y[\mathbf{x}, e_i]$. The relevance of that information regarding the variable of interest y can reasonably be

assumed to depend on the dissimilarity, measured by a suitable distance function, between input vectors $\mathbf{x}$ and $\mathbf{x}_i$. If $\mathbf{x}$ is "close" to $\mathbf{x}_i$ according to the distance function, y is expected to be close to y_i, which makes example e_i quite relevant to predict the value of y. On the contrary, if $\mathbf{x}$ and $\mathbf{x}_i$ are very dissimilar, example e_i provides only marginal information regarding the value of y. Therefore, neighborhood evidence input elements are discounted as a function of their distance to $\mathbf{x}$. $m_y[\mathbf{x}, e_i]$ is defined as a discounting of m_i:

$$m_y[\mathbf{x}, e_i](A) = \begin{cases} m_i(A)\phi(\|\mathbf{x}-\mathbf{x}_i\|), & \text{if } A \in \zeta(m_i) \setminus \{\zeta\} \\ 1-\phi(\|\mathbf{x}-\mathbf{x}_i\|), & \text{if } A = \zeta \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

where ϕ is a decreasing function from $\mathbb{R}^+$ to [0,1].

$$\|\mathbf{x}-\mathbf{x}_i\| = [(\mathbf{x}-\mathbf{x}_i)^T \Sigma^{-1} (\mathbf{x}-\mathbf{x}_i)]^{1/2} \tag{11}$$

where Σ is a symmetric positive definite matrix, a natural choice for ϕ is:

$$\phi(\|\mathbf{x}-\mathbf{x}_i\|) = \gamma \exp(-\|\mathbf{x}-\mathbf{x}_i\|^2) \tag{12}$$

where $\gamma \in [0, 1]$ is a tuning parameter (usually $\geqslant 0.9$).

(3) Pool FBA and deduce prediction output $\hat{y}$.
After acquiring the discounted FBA's m_i, the next is to combine the information provided by each element of the training set using the conjunctive rule of combination of FBA's. The final belief assignment (BA) is then:

$$m_y[\mathbf{x}, \zeta] = \bigoplus_{i=1}^{N} m_y[\mathbf{x}, e_i] \tag{13}$$

and then we normalize FBA to obtain $m_y^*[\mathbf{x}, \zeta]$.

In addition, when the number of BA's combination is increased, the number of focal elements of $m_y[\mathbf{x}, \zeta]$ will increase exponentially and make the computation very heavy for large N. A remedy is using the k nearest neighbors $\{\mathbf{x}_{(i)}\}_{i=1}^{k}$ of $\mathbf{x}$ in the training set to reduce the complexity of the calculation with little loss of accuracy.

3.4.2. Least squares support vector machines (LS-SVM)

SVM is a very nice framework or methodology to formulate the mathematical program for the training error function used in any application. An important appeal of SVM over other traditional regression methods is its ability to handle very high nonlinearity. Similar to nonlinear regression, SVM transforms the low dimensional nonlinear input data space into high-dimensional linear feature space through a nonlinear mapping $\varphi: R^n \to R^{n_h}$, n is the dimension of data space, and n_h is the (very high and even infinite) dimension of the unknown feature space. Then linear function estimation over the feature space can be performed.

LS-SVM (Suykens, Gestel, Brabanter, Moor, & Vandewalle, 2002; Vong, Wong, & Li, 2006) is a variant of SVM with employing least squares error in the training error function, which leads to solving a set of linear equations that is easier to use/solve than quadratic programming (QP) problems, while most of the important advantages of SVM are retained.

Consider the data set $D = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$, with N data points where $\mathbf{x}_k \in R^n, y \in R$, $k = 1$ to N. LS-SVM deals with the following optimization problem in the primal weight space

$$\begin{bmatrix} \min\limits_{\mathbf{w},b,\mathbf{e}} & J_P(\mathbf{w}, \mathbf{e}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \gamma\frac{1}{2}\sum\limits_{k=1}^{N} e_k^2 \\ \text{s.t.} & e_k = y_k - [\mathbf{w}^T\varphi(\mathbf{x}_k) + b], \quad k = 1, \ldots, N \end{bmatrix} \tag{14}$$

where $\mathbf{w} \in R^{n_h}$ is the weight vector of the target function, $\mathbf{e} = [e_1, \ldots, e_N]$ is the residual vector, and $\varphi: R^n \to R^{n_h}$ is a nonlinear mapping, n is the dimension of $\mathbf{x}_k$, and n_h is the dimension of the unknown feature space. Solving the dual of Eq. (14) can avoid the high (and unknown) dimensionality of $\mathbf{w}$. The LS-SVM dual formulation of nonlinear function estimation is then expressed as follows:

$$\begin{bmatrix} \text{Solve in } \boldsymbol{\alpha}, b: \\ \begin{bmatrix} 0 & \mathbf{1}_v^T \\ \mathbf{1}_v & \boldsymbol{\Omega} + (1/\gamma)\mathbf{I}_N \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \end{bmatrix} \tag{15}$$

where $\mathbf{I}_N$ is an N-dimensional identity matrix, $\mathbf{y} = [y_1, \ldots, y_N]^T$, $\mathbf{1}_v$ is an $(N-1)$-dimensional vector $= [1, \ldots, 1]^T$, $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_N]^T$, and $\gamma \in R$ is a scalar for regularization (which is a hyperparameter for tuning). The kernel trick is employed as follows:

$$\boldsymbol{\Omega}_{k,l} = \varphi(\mathbf{x}_k)^T \varphi(x_l) = K(x_k, x_l), \quad k, l = 1, \ldots, N \tag{16}$$

where K is a predefined kernel function.

The resulting LS-SVM model for function estimation becomes

$$\begin{aligned} M(\mathbf{x}) &= \sum_{k=1}^{N} \alpha_k \varphi(\mathbf{x}_k)^T \varphi(\mathbf{x}) + b = \sum_{k=1}^{N} \alpha_k K(\mathbf{x}_k, \mathbf{x}) + b \\ &= \sum_{k=1}^{N} \alpha_k \exp\left(-\frac{\|\mathbf{x}_k - \mathbf{x}\|^2}{\sigma^2}\right) + b \end{aligned} \tag{17}$$

where $\alpha_k, b \in R$ are the solutions of Eq. (15), $\mathbf{x}_k$ is training data, $\mathbf{x}$ is the new input case, and radial basis function (RBF) is chosen as the kernel function K.

4. Experiment verification on methane compressor

Methane compressor is important equipment used in petrochemical industry where normal production flow is required to maintain. Due to the importance of the methane compressor in petrochemical industry, it is imperative to carry out long-term condition monitoring and prediction of future degradation trend in machine health. Therefore, the researches discussed in this paper are vital technologies in this industry field.

In this section, a whole experiment based on the proposed system is described. The experiment object is a low methane compressor shown in Fig. 3. This compressor is driven by an induction motor of 440 kW, 6600 V, two poles with operating speed 3565 rpm. The related information of the machine is summarized in Table 2.

4.1. Experimental setup

The tested system consists of two types of condition monitoring, namely off-line and on-line system. In off-line system, several vibration sensors are installed on selected locations of the motor and compressor, such as drive-end motor (DE), non drive-end motor (NDE), male rotor compressor and suction part of compressor. Each location consists of three directions of measurement: axial, vertical and horizontal. The circle shows the male component of rotor compressor and normal sensing location of this system.

On-line monitoring system consists of acceleration sensors located in only the horizontal direction of four locations, namely, drive-end motor, non drive-end motor, male rotor compressor and suction part of compressor. In addition, a ground-coupling cable was utilized to filter the influence of electromagnetism-noise to the collected signal. The collected signals were amplified through an amplifier and then were sent into a B&K signal analyzer for A/D transform and filtering. At last, the transformed signal was recorded into a desktop for monitoring and analysis.

In this research, a record of the vibration signals of the methane compressor from the date 2005-08-15 to the date 2005-11-22 was analyzed, predicted and validated. The sampling period of the orig-

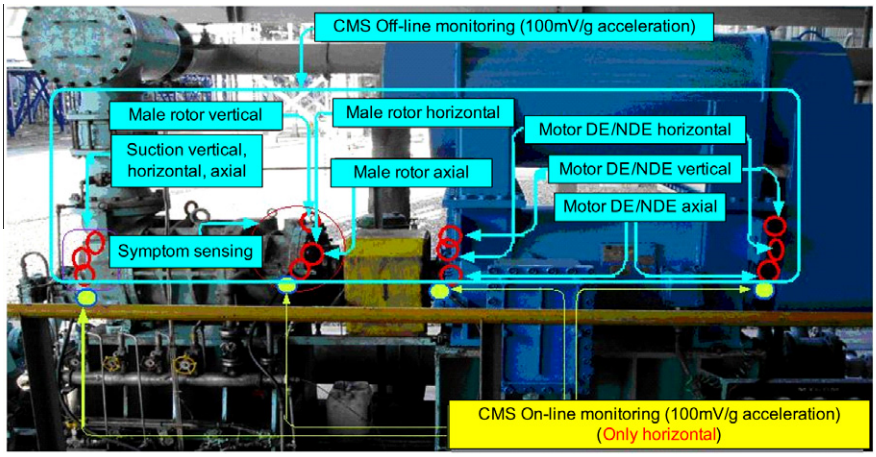


Fig. 3. Low methane compressor: wet screw type.

Table 2
Description of experiment object.

Induction motor		Compressor	
Voltage	6600 V	Type	Wet screw (unload system)
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

inally measured signals was 6 h and four times measurements per day. Thus a total of 400 time-series samples were obtained which involved a process of performance degradation from normal to abnormal running condition.

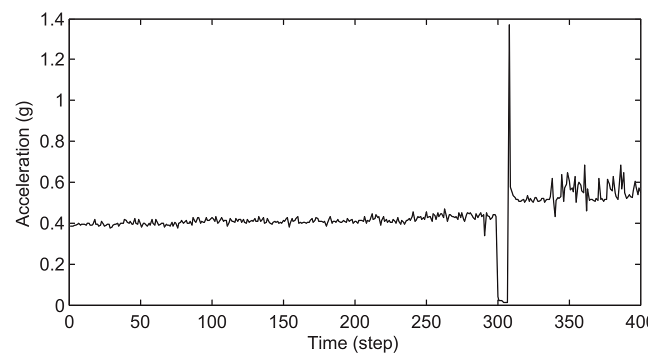
4.2. Experimental results and analysis

4.2.1. Condition monitoring (CM)

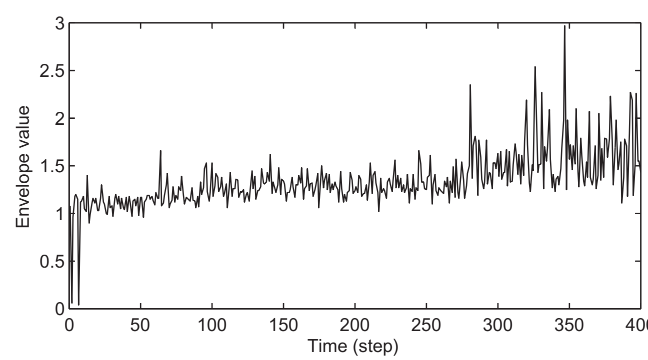
Vibration features, RMS and envelope, are extracted from collected raw signals. RMS is a common feature, even the only indicator used in ISO 10816 and ISO 7919 for machinery condition monitoring and alarming. Envelope is useful to detect glitches (narrow pulse signals). Both of the two features are often used for condition monitoring of rotating machinery. Fig. 4 shows the RMS and envelope plot of peak acceleration data. It can be seen that the collected time-series dataset involves a process of performance degradation from healthy to abnormal operating sate. Both of the features amplitudes show an ascend trend.

After extracting features of vibration signals, a process of normalization is conducted to transform values of features into a common scale and group them as input set for feature-fusion. Next, SOM-based neural network, explained in Section 2.1.1, is employed to combine the input set into a single out indicator, MQE, as shown in Fig. 5. The correlated training parameters of SOM are listed in Table 3. Comparing with the plots of RMS and envelope features, it can be seen that MQE indicator maintains a more steady state than envelop curve, whilst enhances a degradation trend than RMS curve, which is especially appropriate for initial fault detection and health degradation prediction. Therefore, MQE is considered as a good health monitoring indicator for followed analysis.

Furthermore, the step of de-noising is performed to reduce the influence of white noise generated in the just processes of feature extraction and fusion, which is contributable in following trend analysis. A smoothing process of five points moving average is exerted and shown in Fig. 6(a). The smoothness is improved comparing with Fig. 5, but still includes some spines, which maybe trigger



(a) RMS plot



(b) Envelope plot

Fig. 4. Original time-series RMS and envelope plot.

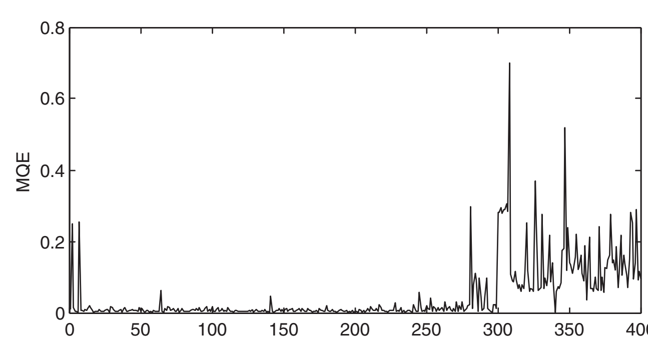
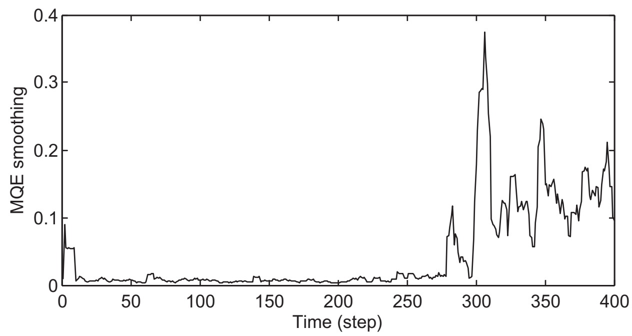
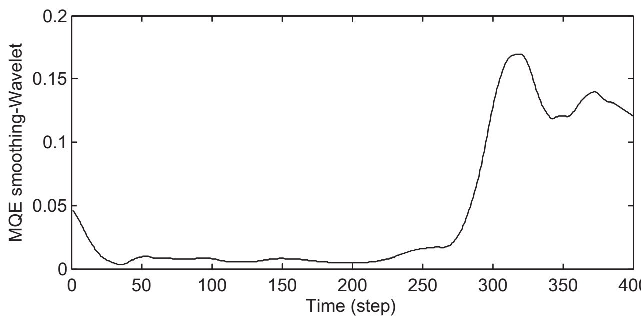


Fig. 5. SOM fusion indicator plot.



(a) Smoothing de-noising



(b) Smoothing-Wavelet de-noising

Fig. 6. Indicator de-noising.

false alarm and make it difficult in making decisions correctly. Therefore wavelet de-noising was conducted in further. The smoothed signal was decomposed at level 5 using 'db5' wavelet function. The processed result indicates a smooth and clear trend as shown in Fig. 6(b).

Based on above de-noised MQE curve, alarm value is set up using the proposed methods in Section 2.3. A data set in healthy and steady operating stage, ranging from point 50 to point 200, was picked out for baseline statistic. The calculated baseline mean value is 0.00696 with the deviation interval [0.00673, 0.00719]. According to the operating history of this machine, the largest constant time was chosen as 20 min. And the span of time window is set as three times of largest constant time. Referencing the calculated values *m*, *k* and Table 1, the alarm coefficient was selected as 4. As a result, alarm mean value was calculated as 0.02784 with the deviation interval [0.02692, 0.02876] referring Eq. (9). As convenience, here only alarm mean value is referenced in this experiment. In addition, we set the trip mean value as six times of alarm mean value in this experiment, which was based on the estimated maximum vibration to which the machine may be subjected.

With the continuous condition monitoring to this machine, the indicator MQE shows an ascending trend as shown in Fig. 7. At the time point of 277, the state curve crosses through the given threshold of alarm. This action triggers a prognostics module and declares a start of performance degradation. Initial fault should be detected; mean while, prediction and evaluation of RUL together with its uncertainty interval of this machine should be carried out.

4.2.2. Degradation prediction and prognostics assessment (PA)

The missions of data-driven prognostics are predicting degradation trend of the machine health, then assessing RUL and its uncertainty interval. For this purposes, two nonlinear regression models, DSR and LS-SVM, are employed.

The step of time-series reconstruction is first exerted using the method of delays (MOD) (Sauer, Yorke, & Casdagli, 1991) for the requirement of nonlinear time-series prediction. In this experiment, the reconstruction parameters, delay time and embedding dimension, are selected by C-C method (Kim, Eykholt, & Salas, 1999) and false nearest neighbor method (Kennel, 1992), respectively. The delay time is chosen as 61 and embedding dimension is chosen as 4.

Next, time-series prediction is performed using DSR and LS-SVM models respectively by strategy of iterated multi-step-ahead (MA). According to the time point of alarm triggered, we divide the previous 277 samples for training models and the left 123 samples for validating the degradation prediction. The relevant training parameters are listed in Table 3. The predicted degradation curve and its prediction interval are shown in Fig. 8 for DSR model and in Fig. 9 for LS-SVM model. The predicted results are compared to the two models in terms of prediction point estimate and prediction interval estimate.

4.2.2.1. Accuracy: prediction point estimate (PP). Accuracy is a measure of how close a point estimate of failure time is to the actual failure time. For degradation prediction, if the prediction is earlier than the actual curve, a correct pre-maintenance (PM) can be carried out. On the contrary, if the prediction is made after actual

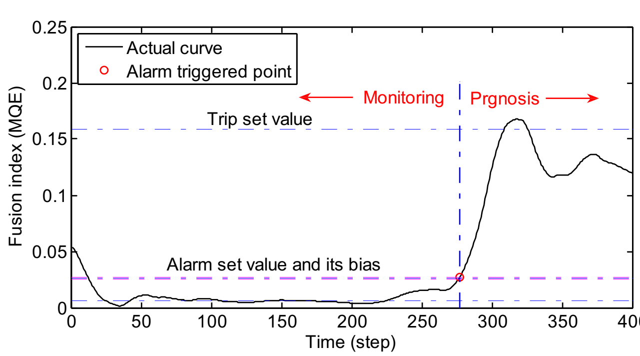


Fig. 7. Condition monitoring and alarming.

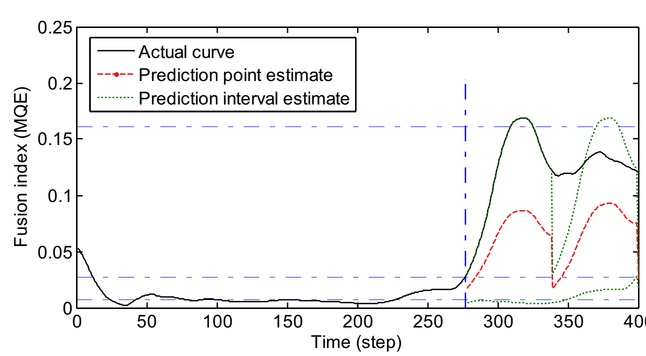


Fig. 8. Degradation prediction using DSR model.

Table 3
List of parameters of models.

Model	SOM	DSR	LS-SVM
Parameters setup	Number of neurons = 5 × 6, Epochs = 10	Initial gamma = 0.9, α = 0.9, *k* = 5	RBF kernel function, initial gamma = 10

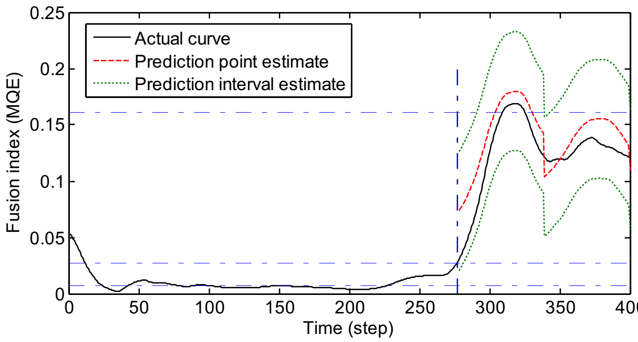


Fig. 9. Degradation prediction using LS-SVM model.

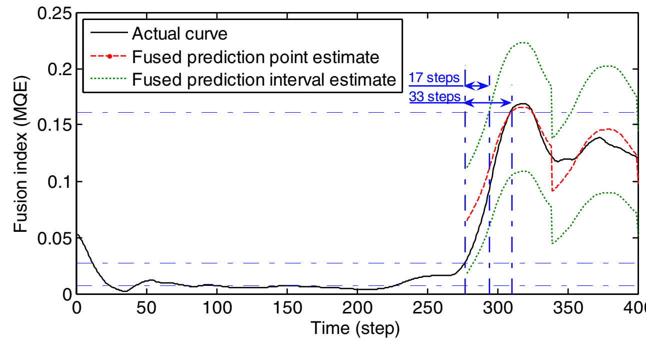


Fig. 10. Degradation prediction fusing DSR and LS-SVM (linear-weight).

failure occurred, prediction becomes meaningless. Therefore, given the same error size, it is in most situations preferable to have a positive bias (early prediction), rather than a negative one (late prediction). Comparing PP estimates in Figs. 8 and 9, DSR could not trigger the trip setting, even though the actual failure occurs. While LS-SVM generates a positive prediction, so it is welcome. Quantitatively, the root mean squared error (RMSE) of PP estimate for LS-SVM is 0.019, less than 0.063 for DSR. Therefore, LS-SVM shows higher accuracy than DSR.

$$E_{RMS} = \left(\frac{1}{n}\sum_{t=1}^{n}(y(t) - \hat{y}(t))^2\right)^{1/2} \tag{18}$$

where $y(t)$ indicates the actual de-noised MQE value of degradation curve and its predicted value is expressed as $\hat{y}(t)$.

4.2.2.2. Precision: prediction interval estimate (PI). Precision is a measure of the narrowness of an interval in which the remaining life falls. The interval is enclosed by upper and lower bounds. A narrow PI estimate indicates high precision. Especially for degradation prediction, a positive upper bound is welcome.

$$Narrowness = \frac{1}{n}\sum_{t=1}^{n}|\hat{y}_{\sup}(t) - \hat{y}_{\inf}(t)| \tag{19}$$

where $\hat{y}_{\sup}(t)$ and $\hat{y}_{\inf}(t)$ stand for the upper and lower bound of predicted degradation curve respectively.

Comparing PI estimates in Figs. 8 and 9, LS-SVM shows a positive upper bound, while the upper bound in DSR is a little reluctant. Quantitatively, the narrowness of PI estimate for LS-SVM is 0.105, less than 0.112 for DSR. Therefore, LS-SVM also shows higher precision than DSR.

After obtaining the prediction results from DSR and LS-SVM, we notice the potential improvements of accuracy or precision by fusing the predicted results. Fusion in this process is already not suitable for using complex methods, because the final fused results should be accordance with alarm setting and trip setting on unit and scale. Hence, linear-weight fusion is employed to combine predicted results of the two models. The weights of the DSR and LS-SVM are set up as 0.15 and 0.85, respectively, considering the previous compared results. The better one deserves a higher weight. The final fused results are shown in Fig. 10. It can be seen that the PP estimate is more close to actual trend curve with RMSE of 0.013. The fused PP is improved comparing with single best LS-SVM, with a proper positive upper bound.

According to the fused prediction results, the PP estimate crosses through the predetermined threshold of trip at point 310, and the upper bound crosses through the threshold at point 294. Finally, the predicted RUL is 33 steps (198 h), with maximum uncertain upper bound of 17 steps (102 h). Comparing with the actual RUL with 33 steps (198 h), the results give a positive prediction, with appropriate time in indicting repair to the machine. Hence, the aims of cost-effective maintenance can be acquired. In on-line monitoring and prediction, the prediction results can be updated continuously, and the predicted RUL will be more accurate after receiving more actual information.

5. Conclusion

This paper suggests a novel mechanism for machine condition monitoring and prognostics on the basis of data-fusion strategy. The proposed system is constructed under the architecture of OSA-CBM, which includes a condition monitoring module and a data-driven prognostics module. The relevant techniques used in this system involve fusion condition monitoring, automatic alarm setting, nonlinear prediction models, prognostics fusion and assessment.

Finally, a whole experiment of methane compressor health monitoring and prognostics is conducted to evaluate the effects of the proposed system. The experiment results show that it can monitor machine health condition effectively and improve accuracy and precision of RUL prediction, thus can enhance maintenance performance obviously.

Acknowledgments

This work is supported by the Technology Development Program for Materials and Components (MOCI) and partially supported by the NURI Project in 2007.

References

Basir, O., & Yuan, X. H. (2007). Engine fault diagnosis based on multi-sensor information fusion using Dempster–Shafer evidence theory. *Information Fusion, 8*(4), 379–386.
Bengtsson, M. (2004). Condition based maintenance system technology – Where is development heading? In *Proceedings of the 17th European maintenance congress, Barcelona, Spain*.
Boutros, T., & Liang, M. (2007). Mechanical fault detection using fuzzy index fusion. *International Journal of Machine Tools and Manufacture, 47*(11), 1702–1714.
Byington, C. S., Watson, M., Kalgren, P., & Safa-Bakhsh, R. (2003). Metrics evaluation and tool development for health and usage monitoring system technology. In *AHS Forum 59, Phoenix, Arizona*.
Chen, A., & Wu, G. S. (2007). Real-time health prognosis and dynamic preventive maintenance policy for equipment under aging Markovian deterioration. *International Journal of Production Research, 45*(15), 3351–3379.
Dong, M., & He, D. (2007). Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis. *Stochastics and Statistics, European Journal of Operational Research, 178*, 858–878.
Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE Transactions on Information Theory, 41*(3), 613–627.
Ginart, A., Barlas, I., Goldin, J., & Dorrity, J. L. (2006). Automated feature selection for embeddable prognostic and health monitoring (PHM) Architectures. In *IEEE systems readiness technology conference, New York* (pp. 195–201).
Goebel, K., & Bonissone, P. (2005). Prognostic information fusion for constant load systems. In *Proceedings of the 7th annual conference on information fusion* (Vol. 2, pp. 1247–1255).

Goebel, K., Bonanni, P., & Eklund, N. (2005). Towards an integrated reasoner for bearings prognostics. *Proceedings of IEEE Aerospace Conference*, 1–11.

Hall, D. L., & Llinas, J. (1997). An introduction to multisensor data fusion. *IEEE Digital Object Identifier, 85*(1), 6–23.

Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing, 20*, 1483–1510.

Kennel, M. B. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A, 45*(6), 3403–3411.

Kim, H. S., Eykholt, R., & Salas, J. D. (1999). Nonlinear dynamics, delay times, and embedding windows. *Physica D, 127*, 4860.

Kohonen, T. (1995). *Self-organizing maps*. Springer.

Lee, J., Qiu, H., Ni, J., & Djurdjanovi, D. (2004). Infortronics technologies and predictive tools for next-generation maintenance systems. In *11th IFAC symposium on information control problems in manufacturing*.

Lee, J., Ni, J., Djurdjanovic, D., Qiu, H., & Liao, H. (2006). Intelligent prognostic tools and e-maintenance. *Computers in Industry, 57*, 476–489.

Nelson, W. B. (2004). *Accelerated testing: Statistical models, test plans, and data analysis* (1st ed.). Wiley-Interscience.

Niu, G., Han, T., Yang, B. S., & Tan, A. C. C. (2007). Multi-agent decision fusion for motor fault diagnosis. *Mechanical Systems and Signal Processing, 21*(3), 1285–1299.

O'Connor, P. D. T. (2002). *Practical reliability engineering* (4th ed.). Wiley-Interscience.

Ogata, K. (1996). *Modern control engineering* (3rd ed.). Prentice Hall.

Petit-Renaud, S., & Denoeux, T. (2004). Nonparametric regression analysis of uncertain and imprecise data using belief functions. *International Journal of Approximate Reasoning, 35*, 128.

Qiu, H., Lee, J., Lin, J., & Yu, G. (2003). Robust performance degradation assessment methods for enhanced rolling element bearing prognostics. *Advanced Engineering Informatics, 17*, 127–140.

Roemer, M. J., & Kacprzynski, G. J. (2001). Assessment of data and knowledge fusion strategies for prognostics and health management. *IEEE Proceedings of Aerospace Conference, 6*, 2979–2988.

Sauer, T., Yorke, J. A., & Casdagli, M. (1991). Embedology. *Journal of Statistical Physics, 65*, 579–616.

Shao, Y., & Nezu, K. (2005). Design of mixture de-noising for detecting faulty bearing signals. *Journal of Sound and Vibration, 282*(3–5), 899–917.

Smets, P. (1998). The transferable belief model for quantified belief representation. In D. M. Gabbay & P. Smets (Eds.). *Handbook of defeasible reasoning and uncertainty management systems* (Vol. 1, pp. 267–301). Dordrecht: Kluwer Academic Publishers.

Smets, P., & Kennes, R. (1994). The transferable belief model. *Artificial Intelligence, 66*, 191–243.

Suykens, J., Gestel, T., Brabanter, J. D., Moor, B. D., & Vandewalle, J. (2002). *Least squares support vector machines*. Singapore: World Scientific.

Thurston, M., & Lebold, M. (2001). *Open standards for condition-based maintenance and prognostic systems, MARCON*. <www.osacbm.org>.

Vong, C. M., Wong, P. K., & Li, Y. P. (2006). Prediction of automotive engine power and torque using least squares support vector machines and Bayesian inference. *Engineering Applications of Artificial Intelligence, 19*, 277–287.

Zhang, J. (2006). Improved on-line process fault diagnosis through information fusion in multiple neural networks. *Computers & Chemical Engineering, 30*(1), 558–571.