Journal of Sound and Vibration 326 (2009) 894–908

Contents lists available at ScienceDirect

Journal of Sound and Vibration

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

A novel method for machine performance degradation assessment based on fixed cycle features test

Linxia Liao *, Jay Lee

NSF I/UCR Center for Intelligent Maintenance Systems, Department of Mechanical Engineering, University of Cincinnati, Cincinnati, OH 45221, USA

ARTICLE INFO

Article history:
Received 16 October 2008
Received in revised form
26 January 2009
Accepted 2 May 2009
Handling Editor: J. Lam
Available online 30 May 2009

ABSTRACT

This paper presents a novel machine performance degradation scheme based on fixed cycle features test (FCFT). Instead of monitoring the machine under constant working load, FCFT introduces a new testing method which obtains data during the transient periods of different working loads. A novel performance assessment method based on those transient data without failure history is proposed. Wavelet packet analysis (WPA) is applied to extract features which capture the dynamic characteristics from the non-stationary vibration data. Principal component analysis (PCA) is used to reduce the dimension of the feature space. Gaussian mixture model (GMM) is utilized to approximate the density distribution of the lower-dimensional feature space which consists of the major principal components. The performance index of the machine is calculated based on the overlap between the distribution of the baseline feature space and that of the testing feature space. Bayesian information criterion (BIC) is used to determine the number of mixtures for the GMM and a density boosting method is applied to achieve better accuracy of the distribution estimation. A case study for a chiller system performance assessment is used as an example to validate the effectiveness of the proposed method.

1. Introduction

Maintenance is usually carried out when failures are serious or have been already occurred in many manufacturing plants, which leads to a plenty of unexpected downtime and correspondingly high maintenance cost. Condition-based maintenance (CBM) has been implemented in most production systems, where certain performance indices are periodically or continuously monitored [1]. The problem is that there is lack of capabilities to convert machinery data into valuable information in order to guide the maintenance practitioners to make right maintenance decisions before the failure happens. For complex or crucial systems, there is no (or rarely) failure mode to follow, which makes it hard to obtain the performance index for the whole system from each monitored parameter. Even if the independent analysis can be carried out for individual parameters, there is always lack of consideration of interactions among components of the whole system. Therefore, robust machine performance assessment methods become a necessity to evaluate of the performance of the entire system and trigger an alarm before serious failure happens.

Some researchers have built physical models for fault detection and diagnosis (FDD) for special systems like chiller systems [2]. Those kinds of methods have been proved effective in some special cases, while it is always not practical when the accurate knowledge for the whole system is inadequate. Moreover, it is not easy to apply the same model to other

* Corresponding author. Tel.: +1 513 556 3820; fax: +1 513 556 4647.
E-mail address: liaol@email.uc.edu (L. Liao).

0022-460X/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsv.2009.05.005

equipment, and usually significant modification to the model is involved. In literature, lots of data-driven methods, which do not incorporate any system knowledge, have been proposed for machine fault detection and diagnosis. For example, logistic regression was used for elevator system degradation detection [3], support vector machine was proposed in [4] to motor system diagnosis, artificial neural network was used in chiller fault detection and diagnosis in [5], and hidden Markov model was considered in [6] to be applied to the tool wear monitoring tasks and so on. To apply those supervised learning methods, measurements from both normal condition and faulty condition are required. Those methods are not feasible when data in the faulty condition is not available, which is usually the case for critical equipment. Some statistical methods [7,8] and unsupervised learning methods like self-organizing maps [9] were proposed for machine faults detection and diagnosis. Those methods have their advantages, but measurements are usually taken when the machine is running only under the fixed working load and failure is usually detected when it has already happened. Some examples using the data under the fixed working load can be found in the literature such as ball bearing race fault detection [10], faults diagnosis for induction motor [11] and chiller faults diagnosis [12] and so on. To find an effective way to detect the incipient faults, wavelet analysis was proposed to detect initial bearing defects in [13] and multivariate statistics were applied to detect incipient defects of gears in [14]. Those methods are developed for specific components and it is hard to apply them for other components or systems. Also, the assumption for normal or nearly normal distribution is necessary for those statistical methods [15], which significantly limits the scope of the applications especially for dynamic processes. Gaussian mixture model (GMM) was proposed in [16] for dynamic process monitoring, while the problem that the expectation–maximization (EM) algorithm may easily converge to a local maximum [17] was not taken care of appropriately in many cases.

This paper proposes a novel method for machine performance degradation assessment based on a fixed cycle features test (FCFT), which identifies the system behaviors and incipient component failure by monitoring the transient periods between different working loads. Only measurement in normal operating condition of the machine is necessary. There is no normal distribution assumption requirement for the proposed performance assessment method. The paper is organized as follows: Section 2 introduces the proposed FCFT test and the method for machine performance degradation assessment. Section 3 illustrates an industrial case of chiller system to validate the effectiveness of the proposed method. Finally, the paper concludes with a discussion of future research in Section 4.

2. Methodology

2.1. A scheme of performance degradation assessment based on fixed cycle features test

The proposed new testing method is called fixed cycle features test. The purpose of FCFT is to identify the system behaviors by monitoring the transient period between different working loads. In general, FCFT includes the following three steps:

(1) Find out the possible spectrum of working loads (such as 100, 75 and 50 percent working load and so on).
(2) In one cycle test, let the system work through all of the possible working conditions with different loads and collect data during the test.
(3) Repeat the test cycles: the timing of various loads in all of the cycles should be approximately identical.

A general scheme of FCFT is illustrated in Fig. 1. FCFT is focusing on the signals in the transient periods during the changes of the working load. Data obtained at three transient periods, which are $T1$ (working load changes from 25 to 50 percent), $T2$ (working load changes from 50 to 75 percent working load) and $T3$ (working load changes from 75 to 100 percent), are used as input signals for the performance assessment models in the next step. A FCFT baseline (under normal operation) needs to be set up first and the performance index is calculated based on the overlap between the recent machine behavior and the baseline. The proposed method for performance degradation assessment based on these data obtained in the FCFT test is shown in Fig. 2.

After the raw data is obtained from the FCFT, signal processing and feature extraction algorithms are used to decompose the multi-sensory data into a feature space which is related to the performance of the machine. Because the FCFT is interested in the signals during the transient period, wavelet packet analysis (WPA), which can capture the dynamic characteristics of the vibration data, is an appropriate algorithm to extract features. If the extracted feature space is in high dimension, principal component analysis (PCA) can be applied to reduce the dimension of the feature space. Gaussian mixture models are then built to approximate the distribution of the feature space. The performance index is evaluated by calculating the overlap between the most recent obtained feature space and the feature space during the normal operation (baseline). This overlap is continuously transformed into a confidence value (CV), following [18], ranging from 0 to 1 (which indicates abnormal and normal, respectively) over time. The CV evaluates the deviation of the recent behavior from the normal behavior or baseline, which is a quantitative measure of the machine degradation. The FCFT is first carried out during the normal condition and the data obtained in this process is used as the baseline. After the baseline is set up, the CV can be calculated to indicate the performance degradation of the machine. Furthermore, a boosting method based on GMM is used to increase the accuracy of density estimation. Bayesian information criterion (BIC) is utilized to find the

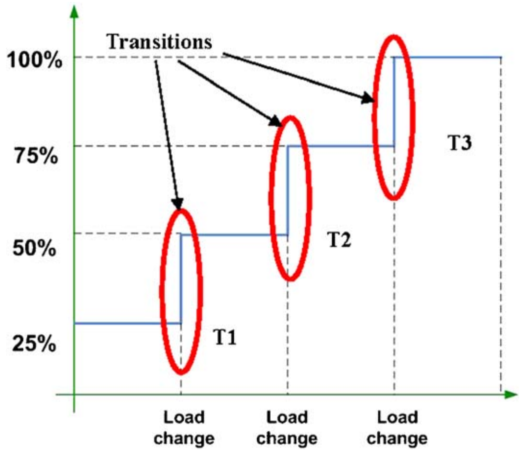

Fig. 1. Fixed cycle features test.

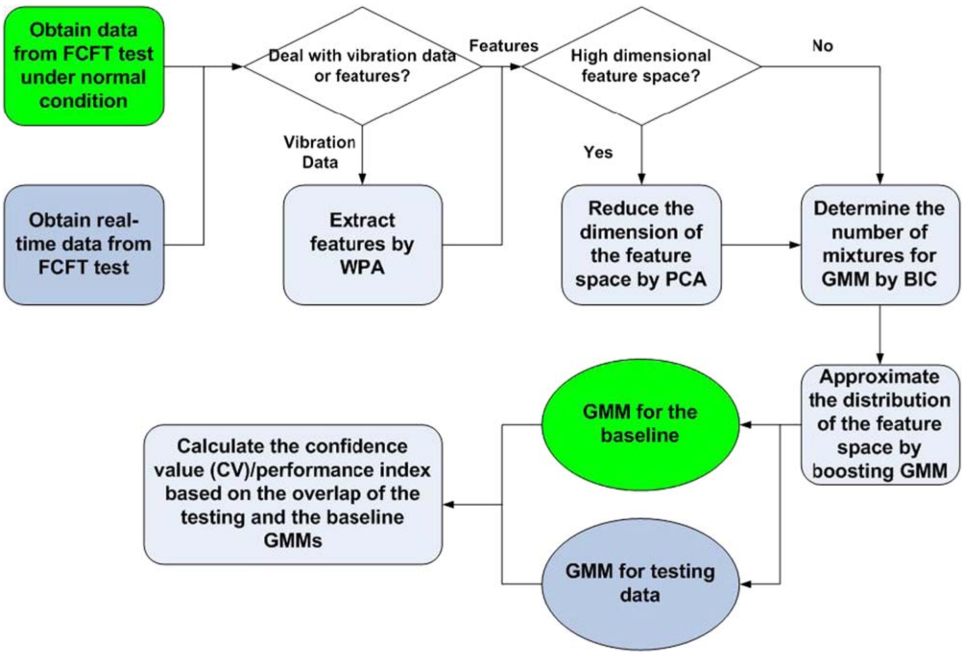

Fig. 2. The performance degradation assessment method based on FCFT.

appropriate number of mixtures for GMM. The following part of this section presents the mathematical background of the aforementioned algorithms to assess the machine performance degradation in detail.

2.2. *Signal processing by wavelet packet analysis and principal component analysis*

2.2.1. *Wavelet packet analysis*

WPA provides a powerful method to deal with non-stationary signal in FCFT. It is an appropriate algorithm for feature extraction in the case, because the vibration signals in the transient period are of our interest. For sustained defects, Fourier-based analysis, which uses sinusoidal functions as base functions, provides an ideal candidate for extraction of these narrow-band signals. For intermittent defects, signals often demonstrate a non-stationary and transient nature. Wavelet packet transform, using a rich library of redundant bases with arbitrary time–frequency resolution, enables the extraction of features from signals that combine non-stationary and stationary characteristics [19]. WPA is an extension of the wavelet transform (WT) which provides complete level-by-level decomposition [20]. The wavelet packets are particular linear combinations of wavelet [21]. The wavelet packets inherit properties such as orthogonality, smoothness and time–frequency localization from their corresponding wavelet functions.

$\Psi^i_{j,k}(t)$ is a wavelet packet function with three integer indices i, j and k which are the modulation or oscillation parameter, the scale parameter, and the translation parameter, respectively.

$$\Psi^i_{j,k}(t) = 2^{j/2}\Psi^i(2^j t - k). \tag{1}$$

When $i = j = k = 0$, $\Psi^0_{0,0}(t) = \varphi(t)$ is called the scaling function. The first wavelet is the so-called mother wavelet (when $i = 1, j = k = 0$, $\Psi^1_{0,0}(t) = \psi(t)$) or analyzing wavelet. In this application, Daubechies wavelet [22] 4 (DB4, shown in Fig. 3), which is a kind of compactly supported wavelets, is used as the mother wavelet.

The following wavelets Ψ^i for $i = 2, 3, \ldots$ are obtained from the following recursive relationships:

$$\Psi^{2i}(t) = \sqrt{2}\sum_{k=-\infty}^{\infty} h(k)\Psi^i(2t - k), \tag{2}$$

$$\Psi^{2i+1}(t) = \sqrt{2}\sum_{k=-\infty}^{\infty} g(k)\Psi^i(2t - k), \tag{3}$$

where $h(k) = 1/\sqrt{2}\langle\varphi(t), \varphi(2t-k)\rangle$ and $g(k) = 1/\sqrt{2}\langle\psi(t), \psi(2t-k)\rangle$ ($\langle\cdot,\cdot\rangle$ stands for the inner product operator) are the quadrature mirror filters (QMF) associated with the predefined scaling function and the mother wavelet function. The wavelet packet coefficients of a signal f can be computed by taking the inner product of the signal and the wavelet packet function:

$$c^i_{j,k} = \langle f, \Psi^i_{j,k}(t)\rangle = \int_{-\infty}^{\infty} f(t)\Psi^i_{j,k}(t)\,\mathrm{d}t. \tag{4}$$

The wavelet packet node energy $e_{j,k}$ is defined as

$$e_{j,k} = \sum_k (c^i_{j,k})^2. \tag{5}$$

The energies of the nodes are used as the input feature space for next step—performance assessment.

2.2.2. Principal component analysis

Principal component analysis is one of the most commonly used statistical methods for reducing dimensionality by transforming the original features into a new set of uncorrelated features. Karhunen–Loève transform (KLT) is a linear dimensionality selection procedure that is related to PCA. The goal is to transform a given dataset $\mathbf{X}$ of dimension $Q1$ to an alternative dataset $\mathbf{Y}$ of smaller dimension $Q2$ in the way that is optimal in a sum-squared error sense [23]. Equivalently, it is seeking to find the matrix $\mathbf{Y}$ which is the Karhunen–Loève transform of matrix $\mathbf{X}$:

$$\mathbf{Y} = \mathbf{A}^{\mathrm{T}}\mathbf{X}, \tag{6}$$

where $\mathbf{A}^{\mathrm{T}}$ is the Karhunen–Loève transform matrix. By choosing the eigenvectors corresponding to the $Q2$ largest eigenvalues of the correlation matrix of $\mathbf{X}$, the mean square error (MSE) between the input $\mathbf{X}$ and its projection $\mathbf{X}'$ is minimized.

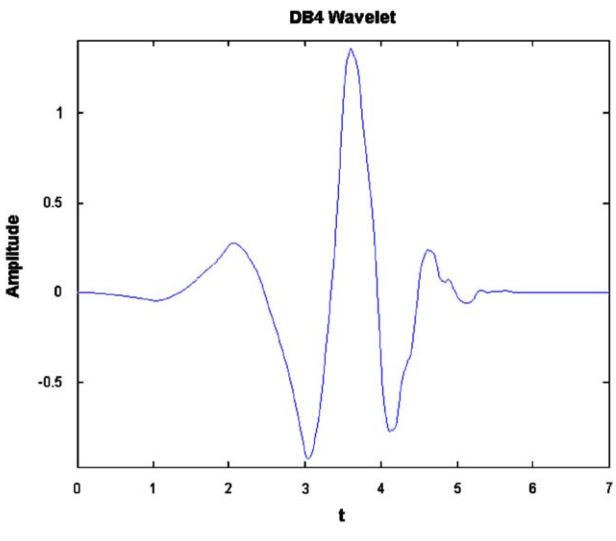

Fig. 3. DB4 wavelet.

2.3. Performance assessment by Gaussian mixture model

Performance assessment by using machine learning algorithms can be separated into two categories (supervised learning and unsupervised learning) by different learning behaviors. In supervised learning, the labels of the data (both normal and abnormal) need to be explicitly used in the model, while the labels of the data are not necessary in unsupervised learning. In FCFT, only the normal condition data (baseline) is available, so an unsupervised learning method should be applied. GMM is an unsupervised learning method which is used to estimate the density distributions of the feature space. GMM consists of a number of Gaussian functions which are combined to provide a multivariate density. Mixtures of Gaussians can be utilized to approximate an arbitrary distribution within an arbitrary accuracy [24]. The mathematic model of GMM is described as

$$f(\bar{x}) = \sum_m p_m Norm(\mu_m, \Sigma_m), \tag{7}$$

where p_m are the weights for the mth mixture and $Norm(\mu_m, \Sigma_m)$ (Eq. (16)) denotes a multivariate Gaussian distribution with mean vector μ_m and covariance matrix Σ_m. If the number of the mixtures is known, expectation–maximization algorithm [23] is usually used to find the proper parameters for the GMM based on the observed dataset.

2.3.1. Determine the number of mixtures by BIC

An obvious problem still remains that is how to choose the appropriate number of mixtures for the GMM. Bayesian model comparison calculates the posterior probabilities by using the full information over the priors. The evidence for a particular hypothesis is calculated by

$$P(D|h_i) = \int p(D|\theta, h_i)p(\theta|D, h_i)\,\mathrm{d}\theta, \tag{8}$$

where θ is defined as the parameters in the candidate model h_i. D represents the training dataset. For globally identifiable cases [25], Eq. (8) can be approximated as

$$P(D|h_i) \approx P(D|\hat{\theta}, h_i)p(\hat{\theta}|h_i)\Delta\theta, \tag{9}$$

where $\hat{\theta}$ is the optimal parameter assumed to maximize $p(\theta|D, h_i)$, $P(D|\hat{\theta}, h_i)$ is the best-fit likelihood and $p(\hat{\theta}|h_i)\Delta\theta$ is the Occam factor [23]. If θ is k-dimensional and the posterior can be assumed to be a Gaussian, then the Occam factor can be approximated directly and yields [23]

$$P(D|h_i) \approx P(D|\hat{\theta}, h_i)p(\hat{\theta}|h_i)(2\pi)^{k/2}|H|^{-1/2}, \tag{10}$$

where

$$H = \frac{\partial^2 \ln p(\theta|D, h_i)}{\partial\theta^2} \tag{11}$$

is a Hessian matrix and measures how "peaked" the posterior is around the value $\hat{\theta}$.

Bayesian information criterion [26] defines a log likelihood function and a penalty term as a criterion for model selection. The BIC score can be calculated by

$$BIC(h_i|D) = \log P(D|\hat{\theta}, h_i) - \frac{d}{2}\log N, \tag{12}$$

where d represents the number of parameters in h_i and N is the size of dataset. For large N, the BIC agrees with the leading order terms in the logarithm of the evidence ($P(D|h_i)$) and so in this case it is equivalent to the Bayesian approach using equal priors for all of the $P(h_i)$ [27] ($P(h_i)$ is a subjective prior for model h_i). The candidate model which has the largest BIC score will be selected as the best model.

2.3.2. Density boosting of GMM

Furthermore, a boosting method based on GMM is used to approximate the density distribution with higher accuracy. Boosting is an algorithm aiming at improving the accuracy of any given learning algorithm or classifiers. In boosting, a weak learner with accuracy on the training set just greater than random guess is first created, and then new component classifiers are added to form an ensemble with high accuracy on the training set by a weighted decision rule. Freund and Shapire's Adaboost algorithm [28] proposed a method to continuously add weak learners until some desired low training error is achieved. In Adaboost, each training pattern is assigned a weight which determines the probability of being selected. If the training pattern is correctly classified, the chance of being selected in the subsequence component classifier is reduced. If the training pattern is not correctly classified, the chance of being selected in the subsequence component classifier is increased. Patterns are chosen according to the new distribution to train the next classifier and the process is iterated. One issue in Adaboost is that the training error is dependent on the labels of the training patterns, while in this case it is unsupervised learning in which the labels are not available. A gradient boosting methodology to the unsupervised learning problem of density estimation method was proposed in [29]. The main idea is to identify the coefficients and parameters of the weak learner which gives the largest local improvement at each iteration step in the data log likelihood (DLL) criterion which is defined as

$$DLL = \log \sum_{t=1}^{T} \alpha_t h_t(x), \tag{13}$$

where T is the number of weak learners, x is the training dataset and α_t is the coefficient for each weak learner $h_t(x)$. In this case, BIC is used as a criterion to choose the number of mixtures for weak learners, which is described in Section 2.3.1. In [30], a GMM boosting method was proposed, but how to select appropriate number of mixtures was not well addressed for unknown structure of data. Another boosting GMM was introduced in [31], in which BIC was used to determine the number of mixtures for the GMM model. However, the number of mixtures should not be defined at the very beginning of the boosting procedure, since the sampled dataset will change according to the weights of the dataset at each iteration step. Besides, the expectation–maximization algorithm, which is utilized to estimate the parameters for GMM, is sensitive to the initial parameters and it probably converges to a local minimum. To address the aforementioned problems, the proposed GMM boosting algorithm is summarized as follows.

Definition of terms in the description of the proposed algorithm:

N	size of the dataset;
i	index of data samples, $1 \le i \le N$;
x_i	ith data sample in the domain of $\mathbf{x}$;
t	current iteration step;
h_t	candidate GMM model at iteration step t;
L_t	boosting GMM model at iteration step t;
α_t	coefficient at iteration step t;
$T_{\max}$	maximum number of iteration steps of the boosting algorithm;
$K_{\max}$	maximum number of iteration steps for the EM algorithm;
$M_{\max}$	maximum number of mixtures of the GMM;
$Q_{\max}$	maximum allowed number of steps if the performance does not improve (log data likelihood changes less than 10^{-5} comparing to the previous step is considered that the performance does not improve);
$\hat{\theta}_{(k-1)}$	estimated parameters of a GMM at iteration step $k-1$. $(k-1)$ in the subscript denotes the parameter(s) at iteration step $k-1$ in the discussion here. The same notation is also used for $\hat{p}_m, \hat{\mu}_m$ and $\hat{\Sigma}_m$, which are the estimated parameters in Eq. (15). E.g. $\hat{p}_{m(1)}$ denotes estimated $\hat{p}_m$ value at iteration step 1. $m = 1, 2, \ldots, M$. M is the number of mixtures and $M \le M_{\max}$.

The sequence of the algorithm is as follows:

1 **Begin** initialize $L_0(x_i)$ to be uniform on the domain of x and set the maximum number of iteration steps $T_{\max}$ and the maximum iteration steps $K_{\max}$ for EM. Set the maximum number of mixtures of the GMM as $M_{\max}$ and set the stop criterion $Q_{\max}$.
2 $t \leftarrow 0$
3 **do** $t \leftarrow t+1$
4 Set

$$w_i = 1/L_{t-1}(x_i) \tag{14}$$

5 Sample the dataset $\mathbf{x}$ according to w_i
6 $M \leftarrow 0$
7 **do** $M \leftarrow M+1$
8 Use EM to estimate the parameters of a GMM model h_t with M mixtures by the sampled dataset $\mathbf{x}$,

where

$$h_t = \sum_{m=1}^{M} p_m Norm(\mu_m, \Sigma_m) \tag{15}$$

and

$$Norm(\mu_m, \Sigma_m) = \frac{1}{(2\pi)^{q/2} |\Sigma_m|^{1/2}} \, e^{-1/2 (x-\mu_m)' \Sigma_m^{-1} (x-\mu_m)}, \tag{16}$$

where q is the dimension of the sampled dataset.

(i) $k \leftarrow 0$, initialize parameters $\hat{p}_{m(0)}, \hat{\mu}_{m(0)}$ and $\hat{\Sigma}_{m(0)}$ for all mixture components

(ii) **do** $k \leftarrow k+1$

(iii)

$$\hat{p}_{m(k)} = \frac{1}{N}\sum_{i=1}^{N}\hat{p}(\omega_m|x_i,\hat{\theta}_{(k-1)}), \tag{17}$$

where ω_m denotes mixture component m, $m = 1,2,\ldots,M$.

$$\hat{\mu}_{m(k)} = \frac{\sum_{i=1}^{N}\hat{p}(\omega_m|x_i,\hat{\theta}_{(k-1)})x_i}{\sum_{i=1}^{N}\hat{p}(\omega_m|x_i,\hat{\theta}_{(k-1)})} \tag{18}$$

$$\hat{\Sigma}_{m(k)} = \frac{\sum_{i=1}^{N}\hat{p}(\omega_m|x_i,\hat{\theta}_{(k-1)})(x_i-\hat{\mu}_{m(k-1)})(x_i-\hat{\mu}_{m(k-1)})'}{\sum_{i=1}^{N}\hat{p}(\omega_m|x_i,\hat{\theta}_{(k-1)})}, \tag{19}$$

where

$$\begin{aligned}\hat{p}(\omega_m|x_i,\hat{\theta}_{(k-1)}) &= \frac{p(x_i|\omega_m,\hat{\theta}_{m(k-1)})\hat{p}_{m(k-1)}}{\sum_{c=1}^{M}p(x_i|\omega_c,\hat{\theta}_{c(k-1)})\hat{p}_{c(k-1)}}\\ &= \frac{|\hat{\Sigma}_{m(k-1)}|^{-1/2}\exp\left(-\frac{1}{2}(x_i-\hat{\mu}_{m(k-1)})'(\hat{\Sigma}_{m(k-1)})^{-1}(x_i-\hat{\mu}_{m(k-1)})\right)\hat{p}_{m(k-1)}}{\sum_{c=1}^{M}|\hat{\Sigma}_{c(k-1)}|^{-1/2}\exp\left(-\frac{1}{2}(x_i-\hat{\mu}_{c(k-1)})'(\hat{\Sigma}_{c(k-1)})^{-1}(x_i-\hat{\mu}_{c(k-1)})\right)\hat{p}_{c(k-1)}}\end{aligned} \tag{20}$$

(iv) **until** $k = K_{\max}$
until $M = M_{\max}$
9 Use the BIC score to determine the best model h_t
10 If $\sum_i w_i h_t(x_i) < N$ [29] break, N is the size of dataset
11 Using line search method to find

$$\alpha_t = \arg\min_\alpha \sum_i -\log((1-\alpha)L_{t-1}(x_i) + \alpha h_t(x_i)) \tag{21}$$

12 Set

$$L_t = (1-\alpha_t)L_{t-1} + \alpha_t h_t \tag{22}$$

13 **until** t $= T_{\max}$or $\log(L_t(x_i)) - \log(L_{t-1}(x_i)) < 10^{-5}$ for $Q_{\max}$ steps
14 **return** L_t
15 **end**

2.3.3. Confidence value/performance index calculation

After the distributions of both normal condition and degraded condition are approximated by boosting GMM, the confidence value, which indicates the performance of the machine (1—normal, 0—abnormal), will be calculated by the overlap of the distributions following [32]:

$$CV = \frac{\int f_1(\bar{x})f_2(\bar{x})\,d\bar{x}}{\sqrt{\int(f_1(\bar{x}))^2\,d\bar{x}}\sqrt{\int(f_2(\bar{x}))^2\,d\bar{x}}}, \tag{23}$$

where $f1$ and $f2$ are the Gaussian mixture functions. If the two distributions overlap extensively, the confidence value will be near 1, which means the performance of the machine is normal. Otherwise, if the two distributions rarely overlap, the confidence value will be near 0, which means a certain abnormal situation happens.

3. An industrial case study of FCFT

3.1. System setup

Chiller system is a complicated system which contains many components such as compressor, condenser/evaporator, water pump and others. Considering that the working load is subjected to change in different working conditions, the monitoring of overall health status of chiller system is not a trivial task. In the FCFT, the whole test cycle is set to be 2 min. The working load changes each 30 s from 25 to 100 percent at a 25 percent interval as Fig. 1 shows. There are six accelerometers (IMI 623C01) installed on the housing of six bearings (channel 0 to channel 5) on the chiller. Channel 0 and

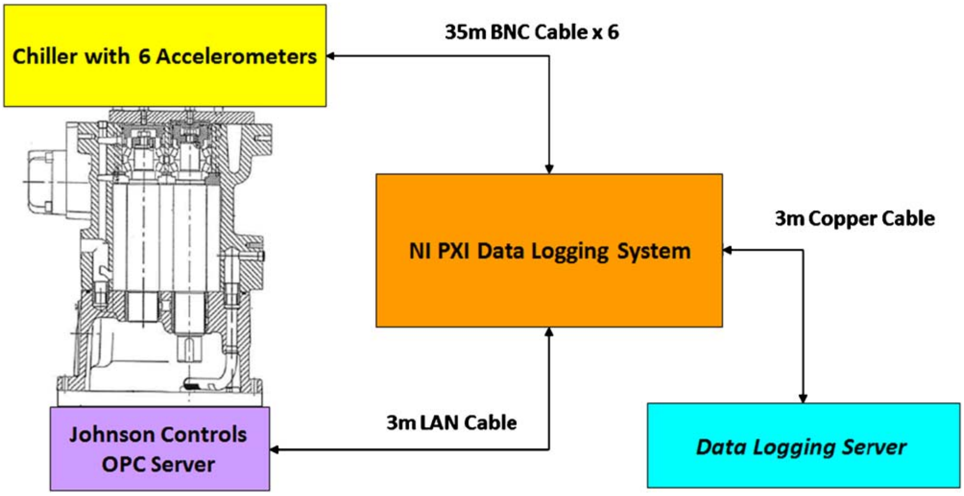

Fig. 4. Hardware setup.

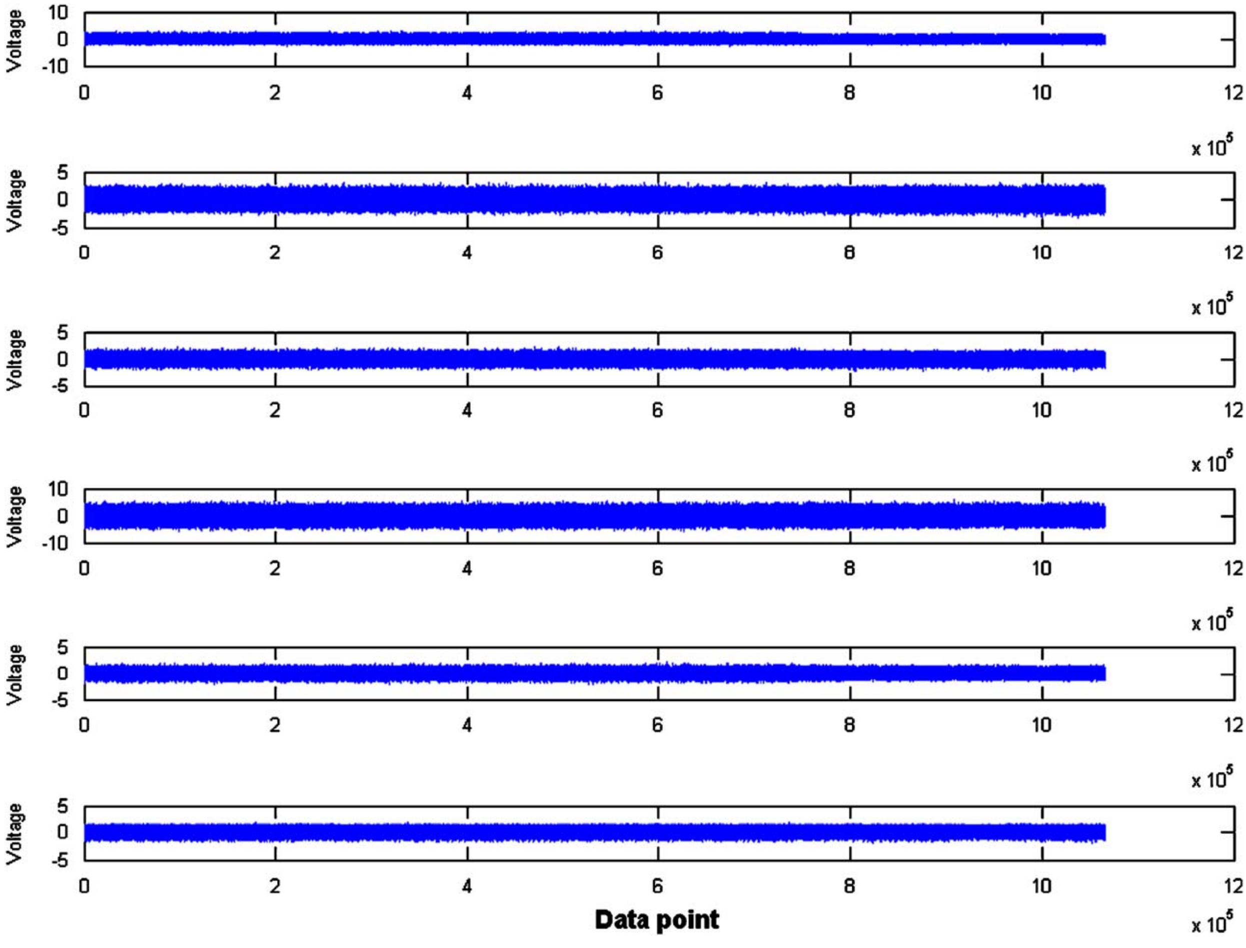

Fig. 5. Vibration data in normal condition.

channel 3 are used for shaft performance monitoring. Channels 1, 2, 4, and 5 are used to monitor bearings ♯1, ♯2, ♯3, and ♯4, respectively.

The structure of the hardware setup is shown in Fig. 4. National Instrument PXI-4472 is used for data acquisition to obtain data from the six accelerometers simultaneously. The sampling rate is set to be 10 K/s for each channel. The data is

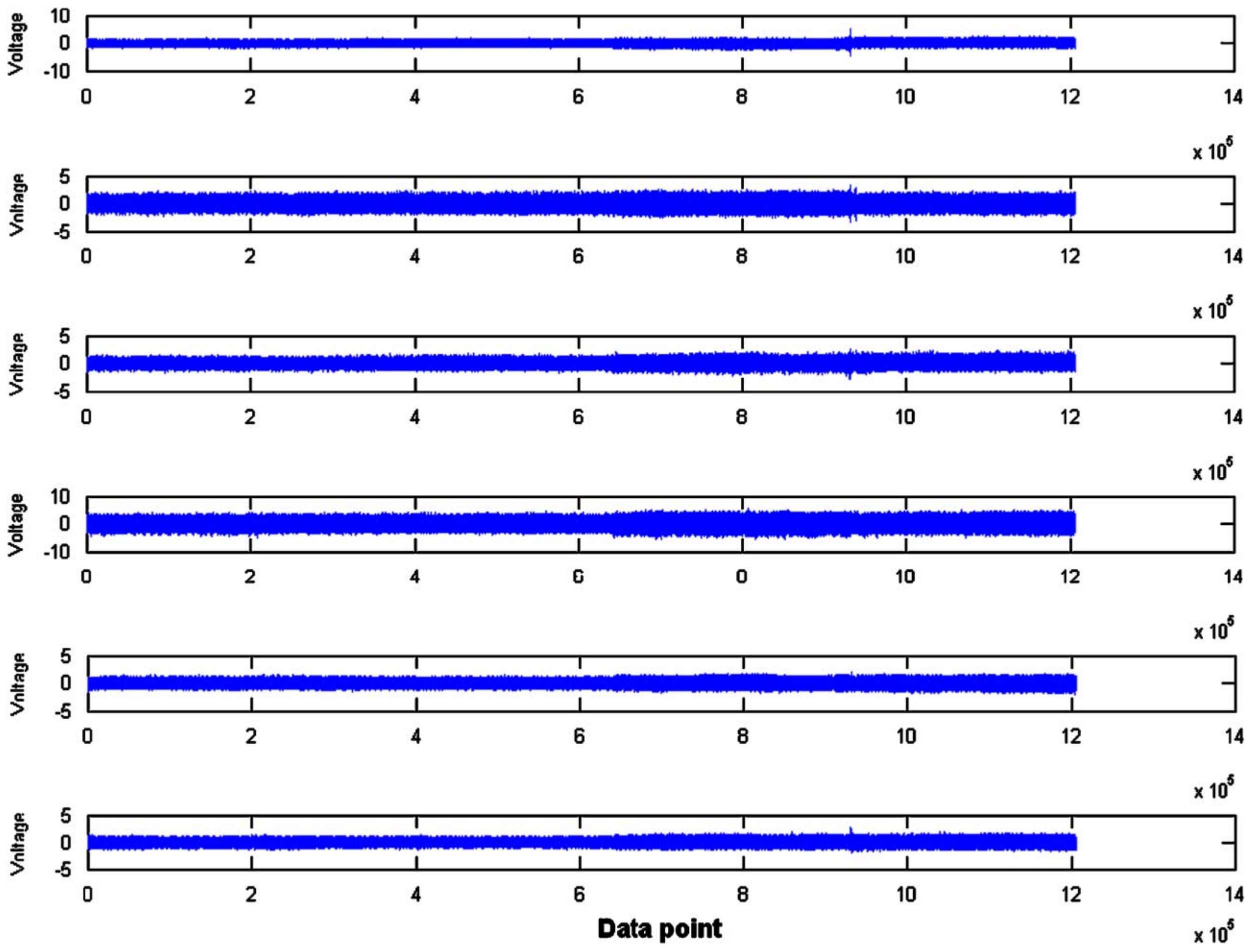

Fig. 6. Vibration data in degraded condition.

Table 1
The OPC parameters and monitored objects.

Monitored object	OPC parameters
Evaporator	Return temperature Supply temperature Flow rate
Condenser	Return temperature Condense supply temperature
Compressor oil	Oil temperature in separator Oil temperature in compressor
Refrigerant circuit	Suction pressure Discharge pressure

obtained for the whole cycle (2 min) of the FCFT. Figs. 5 and 6 show the raw vibration data for the six channels in normal condition and degraded condition, respectively. The degraded condition here refers to an oil leakage problem on a day when the FCFT test was carried out. Field engineers confirmed that it was not a complete failure but an abnormal condition. Obviously, it is not an easy job to identify the performance of each component by just looking at the raw data.

The data logging system also obtains data from the Johnson controls OPC (object-linking—and—embedding for process control) server of the chiller system through Ethernet. The sampling frequency is 4 Hz for each OPC parameter. The monitored objects and the related OPC parameters are listed in Table 1. The OPC parameters in normal condition and degraded condition are illustrated in Figs. 7 and 8, respectively.

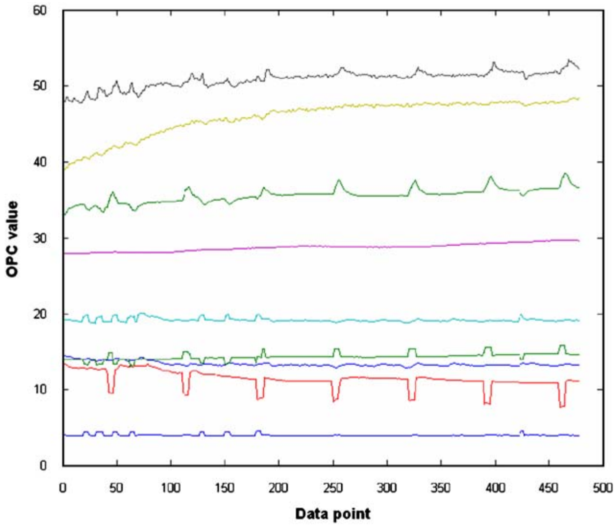

Fig. 7. OPC data in normal condition.

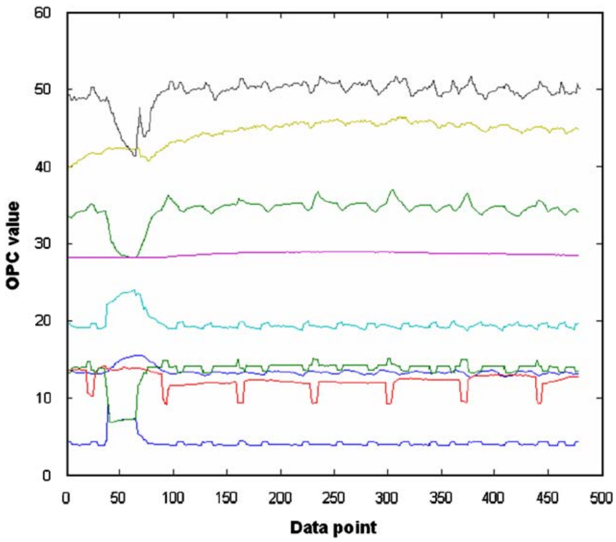

Fig. 8. OPC data in degraded condition.

3.2. Performance assessment results

In the case study, the training dataset and testing dataset of channel 5 (corresponding to bearing ♯4) in normal condition and degraded condition are used as an example to validate the proposed novel method described in Section 2. The normal condition data is analyzed first as the baseline. WPA is used to extract the energy features from the vibration data. PCA is then used to find the first two principal components which contain more than 90 percent of the variation information. Those two principal components are used as the baseline feature space for channel 5 as plotted in Fig. 9.

The next step is using GMM to approximate the density distribution of the feature spaces. The number of mixtures is set from 1 to 10 and the BIC scores of different number of mixtures are calculated by Eq. (12). The results are plotted in Fig. 10 which shows that the BIC score is highest when the number of mixture is 2. Therefore, the number of mixtures is set as 2 to approximate the distribution of the feature space in normal condition.

The approximation results by GMM to the normal feature space are shown in Fig. 11.

After the baseline is set up, the same method can be applied to the newly obtained testing datasets both in normal condition and degraded condition. The BIC score is highest when the number of mixture is 2 in normal condition. A GMM with two mixtures is used to approximate the distribution of the first two principal components of normal feature space. The approximated results combining with that of the baseline are plotted in Fig. 12. The BIC score is highest when the

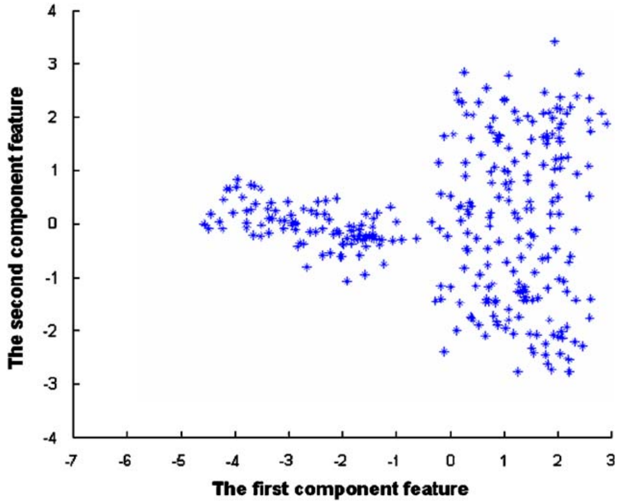

Fig. 9. First two principal components of the normal feature space of channel 5.

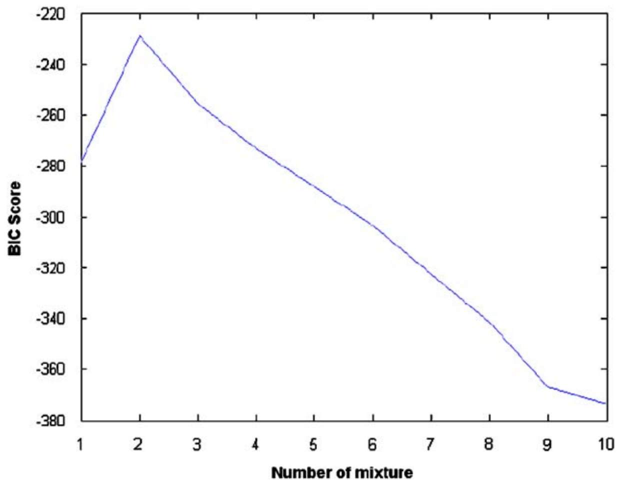

Fig. 10. BIC scores for normal feature space of channel 5.

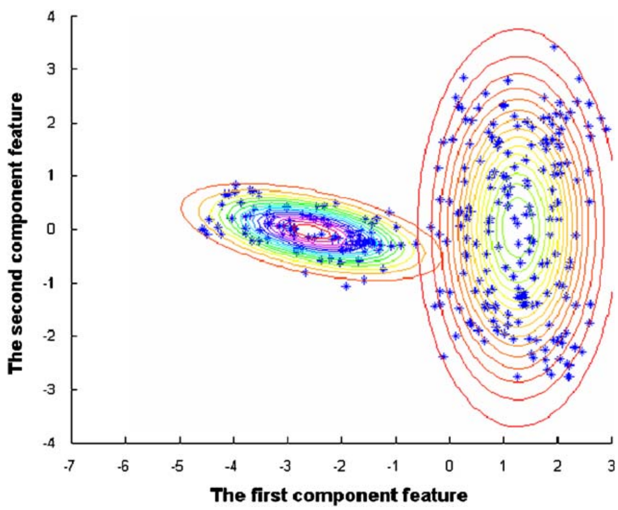

Fig. 11. Results of GMM approximation to the normal feature space of channel 5.

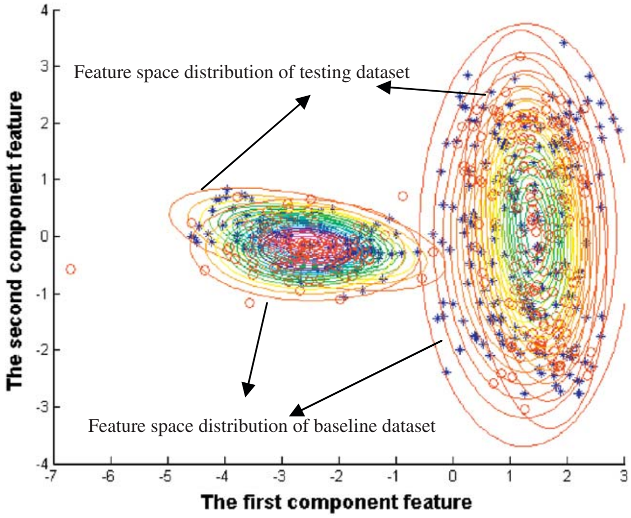

Fig. 12. Results of GMM approximation to the normal and baseline feature space of channel 5.

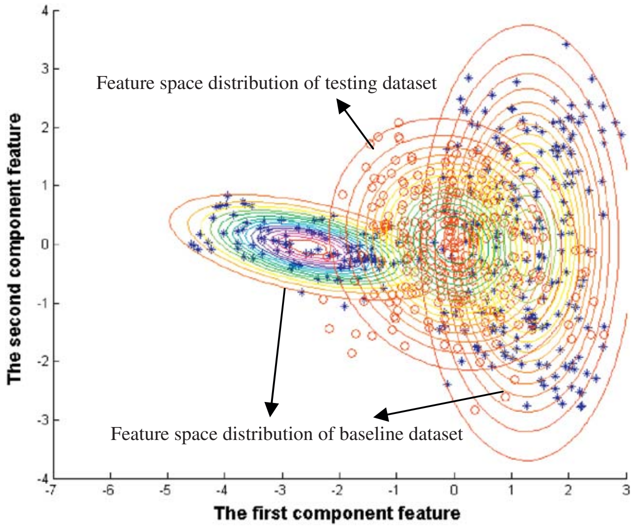

Fig. 13. Results of GMM approximation to the degraded and baseline feature space of channel 5.

number of mixture is 1 in degraded condition. A GMM with one mixture is used to approximate the distribution of the first two principal components of degraded feature space. The approximated results combining with that of the baseline are plotted in Fig. 13.

As shown in Fig. 12, the distributions of the testing feature space and that of the baseline feature space extensively overlap. By using Eq. (23), the calculated CV is 0.9527 which means the performance is normal. As shown in Fig. 13, the distribution of the testing feature space deviates a little from that of the training dataset. By using Eq. (23), the calculated CV is 0.8439 which means the performance is degraded.

The same datasets (baseline and testing data) are also used to validate the proposed density boosting method of GMM. The maximum iteration step for the boosting GMM is set as 25. The maximum number of iteration steps for the EM algorithm and the maximum number of mixtures of the GMM are set to be 50 and 10, respectively. The method using only GMM (without boosting) is also tested on the dataset for 25 times. For each step/time, the data log likelihood is calculated by Eq. (13). Figs. 14 and 15 show the comparison results for the normal condition and the degraded condition, respectively. As the two figures show that no matter it is in normal condition or degraded condition, the boosting GMM method can achieve higher data log likelihood than GMM, which means boosting GMM method represents the distribution of the dataset with higher accuracy. In Fig. 15, the data likelihood achieved by GMM-only varies each time, which means the performance of the boosting GMM is more stable than GMM whose convergence depends on the initial parameters.

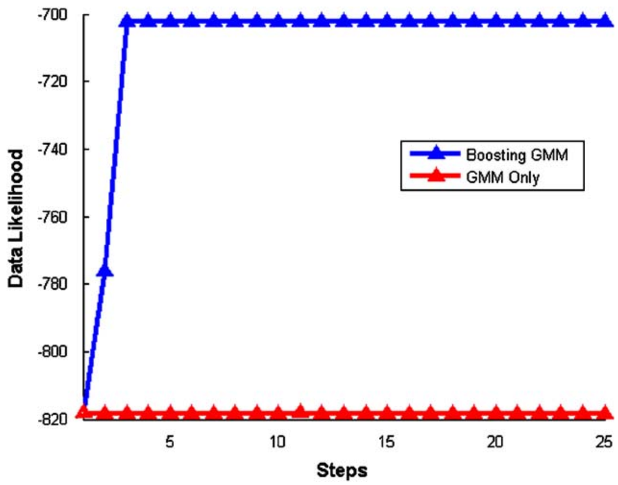

Fig. 14. Comparison of data log likelihood in normal condition for channel 5.

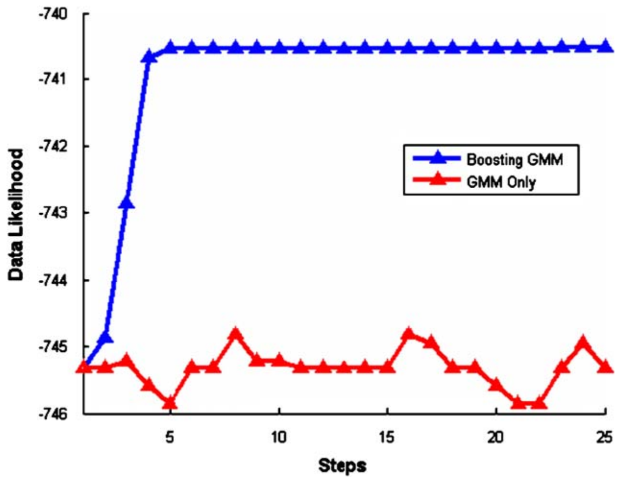

Fig. 15. Comparison of data log likelihood in degraded condition for channel 5.

Table 2
Confidence values in normal and degraded conditions.

	Shaft	Bearing #1	Bearing #2	Bearing #3	Bearing #4
Normal	0.9670	0.9925	0.9781	0.9656	0.9688
Degradation	0.8240	0.8032	0.7289	0.8397	0.8632
	Evaporator	Condenser	Compress oil	Refrigerant circuit	
Normal	0.9969	0.9587	0.9952	0.9882	
Degradation	0.8742	0.8938	0.7502	0.7302	

In this case, the boosting GMM method is applied to all the six channels for both normal condition and degraded condition. Dataset of channel 0 and channel 3 are used to assess the performance of the shaft. The same method is also applied to the OPC dataset without using WPA for signal processing, because the OPC data is considered as features directly. The calculated confidence value for each component is listed in Table 2.

Table 2 presents the confidence values in degraded condition are lower than those in normal condition. After confirmed with the mechanical technicians, the occurrence of the degradation condition is because of the oil leakage problem on the day when the FCFT test was carried out. It is not a complete failure but an abnormal condition. It validated that the proposed machine performance degradation assessment method successfully detected the components' performance deviation from the normal condition before failure happens.

4. Discussion and future research

The proposed FCFT test and machine performance degradation assessment method are effective to identify the system behaviors by monitoring the transient period between different working loads. Actually, from the OPC data shown in Figs. 7 and 8, there is no much difference between the normal condition and the degraded condition when the chiller is running at 100 percent working load (the end part of the data), while the OPC values show difference when the chiller is working at 25 percent working load (the beginning part of the data). If the test is carried out only when the chiller is running at 100 percent working load, nothing abnormal can be detected, while FCFT successfully detect the degradation condition in this industrial case study. Not only for the chiller system, can the proposed method also be potentially applied to other machinery systems.

If the proposed FCFT is carried out every day, a curve of the CV values can be obtained over time. The severe extent of the system can also be evaluated by the trend or decreasing rate of the CV curve. If the system is still acceptable when the CV becomes lower than the preset threshold (e.g. 0.2, which indicates a statistically significant shift from the baseline), the baseline can be updated by using more data when the system is considered as working in normal condition (which depends on the users' tolerance of maintenance). If the system fails one day, besides the normal baseline, the same boosting GMM method can be applied to set up another faulty baseline by using the data when the system is about to fail. Then the CV can be determined by both the normal and faulty baselines. In future research, besides the signal during the transient periods, the stationary signals can also be utilized for component faults diagnosis purposes when the CV drops to an unacceptable level (e.g. 0.2).

Furthermore, using the component feature space as historical data to predict the future feature space is of great research interest. Auto-regressive moving average model (ARMA) [33] and probabilistic neural network [34] are good candidate algorithms for the prediction of the future space. Based on the predicted feature space, the proposed boosting GMM method can be utilized to calculate the predicted CV from the overlap of the predicted feature space and the baseline feature space. This predicted CV which can then be used as a machine performance degradation indicator for preventing failure before it happens.

References

[1] A.K.S. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, *Mechanical Systems and Signal Processing* 20 (7) (2006) 1483–1510.
[2] I.B.D. McIntosh, J.W. Mitchell, W.A. Beckman, Fault detection and diagnosis in chillers—part I: model development and application, *ASHRAE Transactions* 106 (2000) 268–282.
[3] J. Yan, J. Lee, Degradation assessment and fault modes classification using logistic regression, *Journal of Manufacturing Science and Engineering, Transactions of the ASME* 127 (4) (2005) 912–914.
[4] S. Poyhonen, P. Jover, H. Hyotyniemi, Signal processing of vibrations for condition monitoring of an induction motor, *International Symposium on Control, Communications and Signal Processing, ISCCSP*, 2004.
[5] B. Gu, Z.Y. Wang, B.Y. Jing, Fault detection and diagnosis in chillers using artificial neural networks, *Proceedings of the 2003 International Conference on Cryogenics and Refrigeration*, 2003.
[6] L. Atlas, M. Ostendorf, G.D. Bernard, Hidden Markov models for monitoring machining tool-wear, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP*, 2000.
[7] J. Ma, C.J. Li, Detection of localized defects in rolling element bearings via composite hypothesis test, *American Society of Mechanical Engineers, Dynamic Systems and Control Division (Publication) DSC*, 1993.
[8] H. Sohn, K. Worden, C.R. Farrar, Statistical damage classification under changing environmental and operational conditions, *Journal of Intelligent Material Systems and Structures* 13 (9) (2002) 561–574.
[9] L.M.D. Owsley, L.E. Atlas, G.D. Bernard, Self-organizing feature maps and hidden Markov models for machine-tool monitoring, *IEEE Transactions on Signal Processing* 45 (11) (1997) 2787–2798.
[10] S. Prabhakar, A.R. Mohanty, A.S. Sekhar, Application of discrete wavelet transform for detection of ball bearing race faults, *Tribology International* 35 (12) (2002) 793–800.
[11] A. Widodo, B.S. Yang, T. Han, Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors, *Expert Systems with Applications* 32 (2) (2007) 299–312.
[12] S.M. Namburu, J. Luo, M. Azam, K. Choi, K.R. Pattipati, Fault detection, diagnosis and data-driven modeling in HVAC chillers, *Proceedings of SPIE—The International Society for Optical Engineering*, 2005.
[13] H. Qiu, J. Lee, J. Lin, G. Yu, Robust performance degradation assessment methods for enhanced rolling element bearing prognostics, *Advanced Engineering Informatics* 17 (3–4) (2003) 127–140.
[14] N. Baydar, Q. Chen, A. Ball, U. Kruger, Detection of incipient tooth defect in helical gears using multivariate statistics, *Mechanical Systems and Signal Processing* 15 (2) (2001) 303–321.
[15] D. Djurdjanovic, J. Ni, J. Lee, Time–frequency based sensor fusion in the assessment and monitoring of machine performance degradation, *American Society of Mechanical Engineers, Dynamic Systems and Control Division (Publication) DSC*, 2002.
[16] S.W. Choi, J.H. Park, I.B. Lee, Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis, *Computers & Chemical Engineering* 28 (8) (2004) 1377–1387.
[17] K. Krishna, M. Narasimha Murty, Genetic *K*-means algorithm, *IEEE Transactions on Systems, Man, and Cybernetics, Part B:* 29 (3) (1999) 433–439.
[18] J. Lee, Measurement of machine performance degradation using a neural network model, *Computers in Industry* 30 (3) (1996) 193–209.
[19] Z. Sun, C.C. Chang, Structural damage assessment based on wavelet packet transform, *Journal of Structural Engineering* 128 (10) (2002) 1354–1361.
[20] S.K. Goumas, M.E. Zervakis, G.S. Stavrakakis, Classification of washing machines vibration signals using discrete wavelet analysis for feature extraction, *IEEE Transactions on Instrumentation and Measurement* 51 (3) (2002) 497–508.
[21] S. Theodoridis, K. Koutroumbas (Eds.), *Pattern Recognition*, second ed., Academic Press, New York, 2003.

[22] I. Daubechies, Orthonormal bases of compactly supported wavelets, *Communications on Pure and Applied Mathematics* 41 (7) (1988).
[23] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, second ed., Wiley, New York, 2001.
[24] D. Djurdjanovic, J. Lee, J. Ni, Watchdog agent—an infotronics-based prognostics approach for product performance degradation assessment and prediction, *Advanced Engineering Informatics* 17 (3–4) (2003) 109–125.
[25] J.L. Beck, L.S. Katafygiotis, Updating models and their uncertainties. I: Bayesian statistical framework, *ASCE Journal of Engineering Mechanics* 124 (4) (1998) 455–461.
[26] G. Schwarz, Estimating the dimension of a model, *The Annals of Statistics* 6 (2) (1978) 461–464.
[27] J.L. Beck, K.V. Yuen, Model selection using response measurements: Bayesian probabilistic approach, *ASCE Journal of Engineering Mechanics* 130 (2) (2004) 192–203.
[28] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of Computer and System Sciences* 55 (1) (1997) 119–139.
[29] S. Rosset, E. Segal, Boosting density estimation, *Advances in Neural Information Processing* 15, 2002.
[30] X. Song, K. Yang, M. Pavel, Density boosting for Gaussian mixtures, *Neural Information Processing*, 2004, pp. 508–515.
[31] F. Wang, C. Zhang, N. Lu, Boosting GMM and its two applications, *Multiple Classifier Systems*, 2005, pp. 12–21.
[32] G.J. Mclachlan (Ed.), *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, New York, 1992.
[33] J. Liu, J. Ni, D. Djurdjanovic, J. Lee, Performance similarity based method for enhanced prediction of manufacturing process performance, *American Society of Mechanical Engineers, Manufacturing Engineering Division, MED*, 2004.
[34] D.F. Specht, Probabilistic neural networks, *Neural Networks* 3 (1) (1990) 109–118.