Retracted: Manifold learning-based fuzzy $k$-principal curve similarity evaluation for wind turbine condition monitoring

Nanqi Yuan$^{1,2}$ | Wenli Yang$^1$ | Byeong Kang$^1$ | Shuxiang Xu$^1$ | Xiaolin Wang$^2$ | Chengjiang Li$^2$

1Discipline of ICT/School of Technology, Environments and Design, University of Tasmania, Hobart, Tasmania, Australia
2School of Engineering, AMC, University of Tasmania, Hobart, Tasmania, Australia

Correspondence
Nanqi Yuan, School of Engineering, AMC, University of Tasmania, Hobart, Tasmania, Australia.
Email: nanqi.yuan@utas.edu.au

Funding information
China Scholarship Council

Abstract
As wind energy is experiencing an unprecedented development in today’s world, the condition monitoring of wind turbine systems which can avoid serious accidents and economic losses are gathering more and more attentions. Considering the evaluation methods based on machine learning are complicated and unstable in terms of model training and parameter selection, this paper proposed a novel assessment algorithm based on similarity analysis of fuzzy $k$-principal curves (FKPCs) in manifold space. Initially, 38 fusion features containing time-domain, frequency-domain, and wavelet node energy features are extracted from the vibration signal of the wind turbine bearing. Then, the nonlinear local algorithm Laplacian eigenmaps was introduced to transform the original features to the projected lower dimensional space and obtain the more typical parameters. Combining fuzzy clustering with the $k$-principal curve creatively, smooth principal curves of the feature sets were then extracted according to point distribution in three-dimensional manifold space. Finally, the Hausdorff distance algorithm was employed to calculate the similarity between the principal curves of healthy and test datasets to assess the running condition of wind turbine. To validate the present method, the experiments of accelerated bearing degradation were carried out. A series of comprehensive and persuasive comparisons and analysis with different feature conversion methods (principal component analysis [PCA] and Isomap), different distance analysis algorithms (Euclidean distance and Mahalanobis distance), and different assessment models (hidden Markov model [HMM] and deep belief network [DBN]) verified that the proposed FKPC technique can monitor the operating status of the machine more accurately without complex model training and parameter selection processes.

Keywords
condition monitoring, Hausdorff distance, laplacian eigenmaps, manifold learning, principal curve, wind turbine
1 | INTRODUCTION

As the problem of environmental pollution and resource depletion becomes increasingly serious all over the world, wind power, water power, and some other clean energy has been experiencing an ever rapid development, among which wind power that produces no known severe atmospheric pollution has gained a lot of attention. However, since usually being located in remote areas such as the desolate steppes and seacoasts with harsh environment, the relatively high cost of operation and maintenance of wind power systems is hindering its advancement. Wind turbines, the most important parts of wind power generation system, are also the most vulnerable parts due to the complicated internal constitution, especially for the key transmission components such as the gearbox and bearings. However, in practice, online condition monitoring of wind power systems often relies on datasets contaminated with outliers. In view of the large amount of expenses of maintenance when wind turbines break down, running condition monitoring and fault diagnosis under certain circumstances are of tremendous benefit for this issue.

In recent years, traditional periodical maintenance of mechanical equipment has gradually been transforming to condition-based maintenance (CBM). As the prerequisite for a reasonable maintenance strategy, the assessment methods of performance degradation received widespread attention and research interest. Local mean decomposition (LMD) technology, stacked multilevel-denoising autoencoders (SMLDAEs) were all developed for wind turbine gearbox fault diagnosis. And the recently proposed iterative atomic decomposition thresholding (IADT) method which can extract true constituent components of complex signals and suppress background noise interferences also succeeded in fault diagnosis of wind turbine.

As the statistical theory and artificial intelligence are becoming the most widely adopted methods for fault diagnosis and condition monitoring of mechanical equipment, researchers in this field have achieved a wealth of results. A local mean decomposition method, the multimodal deep support vector classification, and the synchronously averaged electric motor signals were applied for gearbox fault diagnosis. The method combining the hidden Markov model (HMM) with Mahalanobis distance successfully reflected the running state of rolling bearings. In addition, a novel prediction method based on least squares support vector machine (LS-SVM) was presented to estimate the degradation trend of a slewing bearing with small sample data. More recently, in a new area of machine learning, deep learning algorithms have successfully been applied in the related fields, too. For example, a novel hierarchical diagnosis network that can collect deep belief networks (DBNs) by layer was proposed for the hierarchical identification of mechanical systems. However, although have seen great developments in both theoretical research and practical applications, statistical methods and artificial intelligence still require a large amount of data for training the mathematical models, and the evaluation results are very sensitive to the parameters of the model. The signal features extracted from mechanical equipment for performance evaluation are usually high-dimensional, nonlinear, information-redundant, and mutual coupling.

Manifold learning is a real low-dimensionality reduction method that can extract the real low-dimensional manifold structure of target data in the original observation space. As one of the classical manifold learning algorithms, Laplacian Eigenmaps (LE) algorithm has been successfully applied to dimensionality reduction and data representation. In contrast, the principal manifold describes the original dataset through constructing an embedded manifold structure (curve or surface) by using a geometric interpretation of principal component analysis (PCA) and approximating the curve or surface to the data distribution and curvature degree of the manifold. The Hausdorff algorithm is generally used to evaluate the similarity among curves, which has been widely studied in the field of object matching and recognition. Also, an improved Hausdorff algorithm with properties of rotation and scale invariance has achieved successful image comparison compared with Mahalanobis distance and Euclidean distance.

In view of the deficiencies about model training and parameter selection of statistics and artificial intelligence-based evaluation techniques, this paper proposed a novel assessment method of analyzing the similarity of fuzzy k-principal curves (FKPCs) in manifold space. The remainder of this paper is organized as follows: Section 2 describes the feature extraction in time domain, frequency domain, and wavelet packet node energies. In addition, using Laplacian Eigenmaps for manifold space projection is explained in detail. In Section 3, the novel FKPC method is proposed and Hausdorff distance is employed to assess the similarity of curves. In Section 4, the experiments comparing various accelerated bearing degradation data among methods like single feature, HMM, and DBN were conducted. Conclusions are presented in Section 5.

2 | FEATURE EXTRACTION AND LAPLACIAN EIGENMAPS

2.1 | Original feature extraction

Vibration signals whose amplitude and frequency change with the performance of bearings are becoming the most commonly used test parameters in bearing experiments. Since it is difficult to extract feature sets that can comprehensively and effectively reflect bearing running state from a single signal analysis method, this paper combines multiple...
methods of time-domain analysis, frequency-domain analysis and wavelet packet transform (WPT) to extract the useful information.

Time-frequency domain features are generally used in the analysis of bearing vibration signals. Time-domain features, which are intuitive and intelligible, constitute the raw data of the bearing running state.\textsuperscript{23} While frequency-domain features are usually used to describe variation in the frequency band in the signal spectrum and spectrum energy distribution. The time-frequency domain features from \( p_1 \) to \( p_{24} \) adopted in this paper for further analysis are shown in Tables 1 and 2.

Wavelet packet transform (WPT) can make an even finer decomposition of the high-frequency portion of the signal without redundancy and oversight, based on decomposition of the low-frequency portion by a discrete wavelet transform. Thus, it can provide better local time-frequency analytic ability than a wavelet transform in a signal process containing middle and high-frequency information about rotating machinery. Wavelet packet node energy is one of the most widely applied features in mechanical fault analysis. The energy value of each node can be solved by Equation (1).

\[
E_{jk} = \int |c_{jk}(t)|^2 \, dt = \sum_{m=1}^{N} x_{jm}
\]

where \( E_{jk} \), \( m \), \( x_{jm} \) are the corresponding signal energy to the reconstructed signal of \( j \)th band in \( k \)th layer, the discrete node and the amplitude of the discrete node respectively.

The feature vector of the wavelet packet can be obtained through normalizing the characteristic parameters calculated by the following formula:

\[
e = \frac{\{E_0, E_1, \ldots, E_f\}}{E}, \quad f = 2^l - 1.
\]

where \( E = \sum_{k=1}^{l} E_{jk} \) is the total energy of the whole signal that equals to the sum of the energy of each subband. After selection, this paper extracted 14 WPT original features for further research.

### Table 1 Features in Time-domain

| Features | Equations | Features | Equations |
|----------|-----------|----------|-----------|
| \( p_1 \) | \( \sum_{n} |x(n)| \) | \( p_7 \) | \( \sqrt{\sum_{n=1}^{N} (\psi_0(n)-p_1)^2} \) |
| \( p_2 \) | \( \sqrt{\sum_{n=1}^{N} (\psi_0(n)-p_1)^2} \) | \( p_8 \) | \( \sqrt{\sum_{n=1}^{N} (\varphi_0(n)-p_1)^2} \) |
| \( p_3 \) | \( \sqrt{\sum_{n=1}^{N} \varphi_0(n)^2} \) | \( p_9 \) | \( \psi_0(n) \) |
| \( p_4 \) | \( \sqrt{\sum_{n=1}^{N} \varphi_0(n)^2} \) | \( p_{10} \) | \( \psi_0(n) \) |
| \( p_5 \) | \( \max \{\psi_0(n)\} \) | \( p_{11} \) | \( \psi_0(n) \) |
| \( p_6 \) | \( \psi_0(n) \) | \( p_{12} \) | \( \psi_0(n) \) |

\( x(n) \) is the sequence of time domain signal, \( n = 1, 2, \ldots N \), \( N \) is the total number of samples.

### Table 2 Features in frequency-domain

| Features | Equations | Features | Equations |
|----------|-----------|----------|-----------|
| \( p_{13} \) | \( \sqrt{\sum_{n=1}^{N} (\psi_0(n)-p_{13})^2} \) | \( p_{19} \) | \( \sqrt{\sum_{n=1}^{N} (\psi_0(n)-p_{13})^2} \) |
| \( p_{14} \) | \( \sqrt{\sum_{n=1}^{N} (\varphi_0(n)-p_{14})^2} \) | \( p_{20} \) | \( \sqrt{\sum_{n=1}^{N} (\varphi_0(n)-p_{14})^2} \) |
| \( p_{15} \) | \( \sqrt{\sum_{n=1}^{N} (\varphi_0(n)-p_{15})^2} \) | \( p_{21} \) | \( \sqrt{\sum_{n=1}^{N} (\varphi_0(n)-p_{15})^2} \) |
| \( p_{16} \) | \( \sqrt{\sum_{n=1}^{N} (\varphi_0(n)-p_{16})^2} \) | \( p_{22} \) | \( \sqrt{\sum_{n=1}^{N} (\varphi_0(n)-p_{16})^2} \) |
| \( p_{17} \) | \( \sqrt{\sum_{n=1}^{N} (\varphi_0(n)-p_{17})^2} \) | \( p_{23} \) | \( \sqrt{\sum_{n=1}^{N} (\varphi_0(n)-p_{17})^2} \) |
| \( p_{18} \) | \( \sqrt{\sum_{n=1}^{N} (\varphi_0(n)-p_{18})^2} \) | \( p_{24} \) | \( \sqrt{\sum_{n=1}^{N} (\varphi_0(n)-p_{18})^2} \) |

\( s(a) \) is signal spectrum, \( a \) is spectral line, \( f_a \) is the frequency of \( a \)th spectral line, \( A \) is the total number of spectral lines, \( a = 1, 2, \ldots A \).

### 2.2 | Laplacian Eigenmaps feature space projection

The original high-dimensional feature extracted in Section 2.1 is usually nonlinear, information-redundant and mutual coupling, leading to dimension disaster and over-fitting, which will increase the algorithm’s space and time complexity. It is thus necessary to generate an essential representation of the original feature matrix by space conversion. The sample data of high-dimensional space (the D dimension) is actually in a low-dimensional manifold (the L dimension), of which the manifold structure is able to retain the geometrical characteristics of the original data. As a classical manifold learning theory, Laplacian Eigenmaps algorithm can obtain a low-dimensional representation of the dataset and can also maintain local features with robust performance when the sample has outliers.\textsuperscript{24}

Given a set \( X = [x_1, \ldots, x_M] \) of M points in \( \mathbb{R}^D \), the purpose of Laplacian eigenmaps is to find an optimal embedding \( \Omega \) for the manifold. Supporting set \( Y = [y_1, \ldots, y_M] \in \mathbb{R}^L \) is low-dimensional representation of \( X \cdot Y = F(X) \), \( F \) represents the embedding map.

A good map \( F \) make the neighboring points in high-dimensional space \( \mathbb{R}^D \) still close to each other in low-dimensional space \( \mathbb{R}^L \) through minimizing the objective function as

\[
\min_{F: \|F(M)\| = 1} \int \left\| \nabla F(X) \right\|^2
\]

\[
= \|F\|_{F^2} = 1.
\]

Note that minimizing Equation (3) can be transformed into a generalized eigenvalue problem:

\[
\Gamma F = -\text{div}\nabla(F).
\]

where \( \Gamma \) is the Laplace Beltrami operator. Let \( f_0, f_1, \ldots, f_M \) be the solutions of Equation (4), ordered according to their eigenvalues \( 0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_M \), therefore, optimal embedding
map can be described as $F = [f_1, f_2, \ldots, f_M]^T$. Then, the high-dimensional space can be projected into low-dimensional manifold space by $Y = F(X) = [f_1(X), f_2(X), \ldots, f_M(X)]^T$.

The main procedures of Laplacian Eigenmaps feature space conversion of rolling machine vibration signals is represented in Figure 1. The extracted time and frequency domain features and WPT features constitute the original high-dimensional (D dimension) feature matrices of the healthy sample and test sample, $M_1^D$ and $M_2^D(t)$ respectively. Then, $M_1^D$ is transformed to the low-dimensional manifold space expressed as matrix $M_1^L$ by Laplacian Eigenmaps, and meanwhile, the project map $F$ is also obtained, which is then employed to calculate $M_2^L(t)$, the low-dimensional representation of $M_2^D(t)$.

3 | SIMILARITY EVALUATION OF FUZZY K-PRINCIPAL CURVES

3.1 | Extraction of fuzzy k-principal curves

The principal curve is a nonlinear generalization of the first principal component, which is a curve that passes through the center of the data distribution.\cite{25} To calculate the principal curve from the dataset is to select a curve satisfying certain optimization objectives and constraints from a set of curves. The principal curve is essentially a one-dimensional manifold embedded in Euclidean space.\cite{26}

Setting the probability density of random variable $Y = (Y_1, Y_2, \ldots, Y_p)$ to $g(s)$, if the curve $f(s)$ through the center of $Y$ data distribution satisfies the following expression

\[
f(s) = E(Y|s_y) = s
\]

then $f(s)$ is one of the principal curves of $Y$, where $s_y = \sup \{s \mid lly - f(s)l = infly - f(s)l\}$ is the projection value from data point $Y$ to point $s$ in curve $f(s)$. The projection of the principal curve is shown in Figure 2.

The $k$-principal curve, also known as the principal curve with length constraint, stresses the length of the constraint curve to ensure good condition rather than abandoning the differentiable conditions of curves.\cite{27} is the only existing proven principal curve. Based on literature,\cite{28} this paper proposes a new method of extracting the principal curve using fuzzy $k$-means clustering. Fuzzy $k$-means is a classical dynamic clustering algorithm that constantly minimizes the square error and updates the membership degree of each data sample point to the cluster center.\cite{29} it is, therefore, the cluster with the highest membership degree is actually the classification of the data sample.

For $N$ samples with $p$ variables after normal transformation and standardization, a fuzzy matrix of the initial $m$ groups of each sample is given as

\[
U = \begin{bmatrix} u_{i1} & \cdots & u_{im} \\ \vdots & \ddots & \vdots \\ u_{n1} & \cdots & u_{nm} \end{bmatrix} \quad i = 1, 2, \ldots, n; j = 1, 2, \ldots, m.
\]

(6)

where $u_{ij} \in [0, 1]$, $\sum u_{ij} = 1$, $u_{ij}$ is the membership degree of sample $i$ for class $j$. Let $v_j$ be the initial cluster center of class $j$

\[
V = [v_1, v_2, \ldots, v_m]
\]

(7)

where $v_j = [v_{j1}, v_{j2}, \ldots, v_{jp}]$, then determining an objective function $J_\zeta(U, V)$

\[
J_\zeta(U, V) = \sum_{i=1}^{n} \sum_{j=1}^{m} (u_{ij})^2 \|x_i - v_j\|^2.
\]

(8)
This problem can be attributed to solve the extreme value of the objective function on the condition of $\sum u_j=1$. When $m>1, x_i \neq v_j$, by using Lagrange multiplication, which can be proved as follows\(^{30}\)

\[
u_j = \frac{1}{\sum_{j=1}^m \left( \frac{\|x_i - v_j\|}{\|x_i - v_1\|} \right)^{2/(c-1)}}
\]

\[
v_j = \frac{\sum_{i=1}^n u_i^j x_i}{\sum_{i=1}^n u_i^j}, \quad i=1,2,\ldots,n; j=1,2,\ldots,m
\]  \tag{9}

The formation of $k$-principal curve is to minimize Equation (10) continuously, then, a 2-opt algorithm is used to optimize the Hamilton path formed by line segments until they are connected into a smooth curve.\(^{31}\)

\[
\min \sum_{i=1}^k \sum_{x \in V_i} d(x, s_i)^2
\]  \tag{10}

where $x$ denotes a point, $s_i$ the $i$th line segment, $d(x, s_i)$ the distance from $x$ to $s_i$, and $k$ the total number of line segments.

On account of the above illumination, the proposed method for extracting principal curves can be summarized as: (a) Using Fuzzy $k$-means clustering algorithm to divide the original data points into $m$ classes; (b) Iterating continuously to minimize Equation (10) and obtaining $k$ line segments; (c) Optimizing the Hamilton path to connect $k$ line segments into a smooth curve.

The contrast effect of the first principal component and the $k$-principal curve is displayed in Figure 3, which shows that the principal component reflects the overall distribution of dataset, while $k$-principal curve is the nonlinear extension of the fitting and principal component with the ability of describing the trend of dataset to the greatest extent.

### 3.2 Curve similarity evaluation

The judgment analysis of curve similarity is one of the foremost research contents of computer graphics and pattern recognition, of which the techniques mainly include the eigenvalue method and the similarity functions definition-based method. The eigenvalue method applies machine learning to compare characteristic parameters, whereas the similarity functions definition-based method analyzes similarity through calculating the distance between two curves, and according to the previous researches the later is a little superior to the former.

As a widely adopted similarity measurement theory between two-point sets, the Hausdorff distance was successfully applied in object retrieval,\(^{20}\) information processing\(^{32}\) and some other fields. The Hausdorff distance is a nonlinear distance, which only calculates the degree of similarity (maximum distance) between two-point sets without establishing the corresponding relationship between the points.\(^{33}\)

For points sets $A = \{a_1, a_2, \ldots, a_p\}$ and $B = \{b_1, b_2, \ldots, b_q\}$ then the Hausdorff distance based on average distance value is

\[
H_{\text{MHD}}(A, B) = \max \{ h_{\text{MHD}}(A, B), h_{\text{MHD}}(B, A) \}. \tag{11}
\]

Where

\[
h_{\text{MHD}}(A, B) = \sum_{a \in A} h(a, B)/p, h_{\text{MHD}}(B, A) = \sum_{b \in B} h(A, b)/q
\]

and

\[
h(a, B) = \max_{a \in A} \min_{b \in B} ||a - b||. \tag{12}
\]

\[
h(A, b) = \max_{b \in B} \min_{a \in A} ||a - b||. \tag{13}
\]

Note that $||a - b||$ is the Euclidean distance between points $a$ and $b$.

### 3.3 Machine performance assessment based on the present FKPC

In this work, the original high-dimensional feature matrix is transformed to the three-dimensional space with the help of Laplacian eigenmaps, then the fuzzy $k$-principal curve $f^0$ of the healthy samples as well as the $f^t(t)$ of the test samples are extracted, respectively, note that $t$ is the running time of the machine. Next, the outcome of machine performance assessment can be calculated through resolving $\text{Dist}(t) = H_{\text{MHD}}(f^0, f^t)$, of which the algorithm procedures are presented in Table 3.

---

**FIGURE 3** The first principal component and the principal curve

---

---
Assessment algorithm of machine running performance

Input: bearing vibration signal

Step 1: Extracting the feature matrix,
Healthy samples,
$$\mathbf{M}^p = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{M1} \end{bmatrix} = \begin{bmatrix} p^1_1 \\ \vdots \\ p^1_{M1} \end{bmatrix}$$
Test samples, $$\mathbf{M}^p(t) = \begin{bmatrix} p^1_1(t) \\ p^1_2(t) \\ \vdots \\ p^1_{M1}(t) \end{bmatrix} = \begin{bmatrix} p^1_1(t) \\ \vdots \\ p^1_{M1}(t) \end{bmatrix}$$

Step 2: Space conversion based on Laplacian eigenmaps
Calculate Equation (4) to obtain the low-dimensional representation of data set $$\mathbf{M}^p$$ and project map F
Calculate $$\mathbf{M}^p_1 = F(\mathbf{M}^p)$$

$$\mathbf{M}^p \rightarrow \mathbf{M}^p_1 = \begin{bmatrix} p^1_1 \\ \vdots \\ p^1_{M1} \end{bmatrix}$$
$$\mathbf{M}^p(t) \rightarrow \mathbf{M}^p_1(t) = \begin{bmatrix} p^1_1(t) \\ \vdots \\ p^1_{M1}(t) \end{bmatrix}$$

Step 3: fuzzy k-principal curve extraction
Calculate the fuzzy k-principal curves of healthy samples $$f^0$$ and test samples $$f^1(t)$$ based on 3.1

Step 4: Similarity assessment, $$\text{Dis}(t) = H_{\text{MHD}}(f^0, f^1(t))$$

TABLE 3  Algorithm procedures of the proposed FKPC

FIGURE 4  The experimental platform
EXPERIMENT ANALYSIS

4.1 Experimental setup

To simulate the scene in which the wind turbine gradually fails day-by-day in practical applications as realistically as possible, the bearing run-to-failure experiment was implemented on the test devices shown in Figure 4. The experimental platform is composed of the rotating part, the degradation generation part (with a radial force applied on the tested bearing) and a measurement part. The rotating part includes the asynchronous motor with a gearbox and its two shafts: the first one is near to the motor and the second one is placed at the ride side of the incremental encoder. Components of the loading part are grouped in a unique and same aluminum plate, which supports a pneumatic jack, a vertical axis and its lever arm, a force sensor, a clamping ring of the test bearing, a support test bearing shaft, two pillow blocks and so on. To obtain the accurate vibration signal data, the test bearing was fixed with two high sensitivity miniature accelerometers positioned at 90° to each other, and the NI DAQ Card 6062E was also applied in the data acquisition system. The parameters of the bearings and operation conditions are listed in Table 4.

In order to avoid propagation of damages to the whole test bed (and for security reasons), experiments were stopped when the amplitude of the vibration signals (horizontal and vertical) overpassed 20 g. Figure 5 depicts an example of the bearing components before and after the test. When the experiment was finished, 985 individual ASCII format data files were got, of which each file consists of 2560 data points with the recording interval of 10 seconds. And the outer ring of selected bearing was found to be faulty after the test.

4.2 Experimental results

After the data preprocessing of the signal data collected from the above experiment, four data files were found to be abnormal and removed, the rest 981 files were adopted for further analysis. For every 256 data points, time and frequency domain analysis and WPT were utilized to extract the 38 features in original feature space, here a 9810 x 38 array composed of original features was obtained. The eight representative original features that have been widely used for further research are displayed in Figure 6, in which all the waveforms of the eight features have an obvious mutation at time point 7000 second, however, only four of them (MF, WPT1, WPT5, and WPT6) show the slight abnormality at about time point 5300 seconds, while other four features (Skewness, Kurtosis, Crest, and SDF) seem unable to detect the early slight abnormality.

Then, the original feature matrix of the healthy bearing state was transformed into three-dimensional space with preserved projection map information by using Laplacian Eigenmaps, while the feature matrix of the test bearing was also transformed to three-dimensional space according to the projection map. Figure 7 shows the feature distributions after space conversion, where the bearing samples can be clearly distinguished.

The fuzzy k-principal curve is extracted from the feature sets after space conversion, and then the similarity between the principal curves of test samples and healthy samples is calculated by using Hausdorff during each time period. Figure 8 is a schematic diagram of the fuzzy k-principal curve of a bearing running for 5000 seconds, in which the red line and blue dots represent fuzzy k-principal curve and original dataset of the bearing running on 5000 seconds respectively.

TABLE 4  Bearing parameters and experimental conditions

| Outside race diameter (mm) | Inside race diameter (mm) | Diameter of balls (mm) | Motor Power (W) | Accelerometers Type | Rotation speed (rpm) | Maximum load (N) | Sampling rate (kHz) |
|---------------------------|---------------------------|------------------------|----------------|---------------------|---------------------|-----------------|-------------------|
| 32                        | 20                        | 3.5                    | 250            | DYTRAN 3035B        | 57.5                | 4000            | 25.6              |
The normalized trend chart for similarity assessment of the principal curve after space conversion is plotted as Figure 9, it can be discovered that no significant variance in the distance appears from 0 to 5100 seconds, which indicates that there is no fault in the early running period and the bearing is under normal condition. However, the distance begins to increase after 5100 seconds, indicating the appearance of subtle wear and a gradual deviation from the normal operating condition. After 7020 seconds, the distance suddenly changes and then falls back with a slight fluctuation, which suggests that the bearing fault deepened and the bearing was gradually polished. While after 9600 seconds, the rapid increase of the distance points out a serious bearing degradation as the bearing reached the failure state. Comparing with the actual experiment situations, it can be discovered that the assessment result of the proposed FKPC method in this study.

**FIGURE 6** Eight of the original features

**FIGURE 7** Feature distribution after Laplacian eigenmaps

**FIGURE 8** Fuzzy $k$-principal curve on 5000 min

**FIGURE 9** Degradation trend curve of bearings
4.3 Comparisons and discussions

4.3.1 Comparisons with different feature conversion methods

The Laplacian Eigenmaps method adopted in this study is a nonlinear local feature space conversion method, in order to make comparative analysis and highlight its effectiveness for the work, several contrast experiments with linear space conversion method PCA and global nonlinear space conversion method Isomap that are proverbially applied for feature transformation were carried out. The feature distributions after dimension reduction of PCA and Isomap are displayed in Figure 10, compared with the result of Laplacian Eigenmaps in Figure 7 one can find that it is more obviously to distinguish the bearing samples by using Laplacian feature mapping and PCA, while Isomap is not able to distinguish the samples effectively.

The $k$-principal curves of the PCA and Isomap features were extracted after feature conversion, and then the principal curve similarity analysis was also carried out through Hausdorff theory during each time period, of which the results is shown in Figure 11. It can be learned that although the methods of Laplacian and PCA can both portray the general degradation trend of bearing performance apparently, the Laplacian features are more stable before 5100 seconds and can reflect the distance variation after 7020 seconds more obviously. While the distance fluctuation without obvious tendency indicates that the Isomap method is not capable for describing the trend of bearing degradation, which may be caused by the fact that Isomap is suitable for the low-dimensional manifolds whose interior is flat but not appropriate to the manifold with larger internal curvature.

4.3.2 Comparisons with different distance analysis methods

In addition to Hausdorff distance, Euclidean distance and Mahalanobis distance are also commonly employed in the analysis of curve similarity. Next, the comparison experiments adopting these two kinds of methods to evaluate the similarity of the $k$-principal curves between test bearing and healthy bearing were carried out. Figure 12A,B are the normalized assessment curves of Euclidean distance and Mahalanobis distance, respectively, from which it can be learned that both Euclidean distance and Mahalanobis distance is able to evaluate the performance degradation of the machine, especially
for the serious faults. However, they only detected the early faults occurred near 6100 and 6400 seconds, respectively, which indicated that both of them are not as sensitive to the early machine failure as Hausdorff distance who successfully detected the early minor faults at about 5100 seconds.

4.3.3 | Comparisons with HMM and DBN

In the existing research achievements, the HMM and DBN algorithms are successfully and widely employed in pattern recognition, data processing, machine state assessment, and some other fields. Typically, HMM uses the normal feature sets to train the model and then uses the test feature sets and the well-trained model to calculate the probability value, during which the greater the deviation away from the healthy state is, the smaller the probability value becomes. While as a probabilistic generative model composed of a series of restricted Boltzmann machines (RBMs), DBN is commonly utilized in combination with a back propagation neural network (BPNN) in the field of mechanical fault analysis. To highlight the advantages of the proposed FKPC method in this paper, the similar comparative experiments with the above two artificial intelligence models were conducted in the following study.

In the comparison experiments, the feature space conversion method remained unchanged as Laplacian Eigenmaps, while the evaluation models were replaced by HMM and DBN, respectively, of which the results after normalization are represented in Figure 13. The inconspicuous tendency of performance degradation and the slight abnormality, which did not appear until 5900 seconds in Figure 13A, reflected the fact that HMM is not sensitive to early machine failure. Furthermore, the curve did not feature the obvious mutational rising fluctuation as seen in Figure 9 at around 7000 seconds either, so it can be considered that HMM cannot accurately describe the running state of a machine from fault deepening to total failure. As described in Figure 13B, DBN can effectively detect the serious degradation such as crackle, fatigue spalling etc. occurred at 7000 seconds, but it failed to detect the early slight faults of wear, pitting, or overheat began in the vicinity of 5100 seconds, and the undulate forepart as well as the disorganized end of the assessment curve also further indicated that DBN is not competent enough for this task. Additionally, the complicated parameters such as the number of hidden states and Gaussian mixture functions of HMM and the number of hidden layers and neurons of DBN, and the erratic training processes of these two models would make the applications of them more difficult.

In order to carry out a more accurate quantitative analysis of the assessment effectiveness of the above three comparative algorithms, the time points when the relative change rate rises to 2.0%, 5.0% and 30.0% of each result curve were calculated as displayed in Table 5. The relative change rates were adopted to represent the early slight
fault, normal fault, and serious fault respectively. It can be learned from the quantitative analysis table that the present FKPC method is the quickest one to achieve the change rate at 2.0%, 5.0%, and 30.0%, while the DBN achieved a little worse and HMM performed worst, which is pretty in line with the analysis results of the trend curves. The above comparison analysis further indicated that the proposed FKPC method in this paper can portray the degradation trend of wind turbine bearing performance more accurately and possesses greater sensitiveness to early slight faults than the HMM and DBN.

5 | CONCLUSIONS

The condition monitoring and performance assessment of wind turbine has important pragmatic significance. This work reports a novel assessment algorithm of wind turbine bearing performance degradation based on fuzzy k-principal curve similarity in manifold space. First, three popular approaches including time and frequency domain analysis as well as WPT were applied to extract 38 features of the vibration signals collected from machines in the original high dimension space. Next, the nonlinear local algorithm Laplacian eigenmaps was introduced to transform the original features to the projected lower dimensional space and obtain the more typical parameters. Then, the FKPCs of the feature sets after space conversion were extracted. Finally, the Hausdorff distance theory was applied for calculating the similarity between the curves of the test samples and healthy samples to analyze the degradation tendency of the wind turbine bearing performance. In order to validate the effectiveness of the proposed algorithm, the bearing run-to-failure experiment was carried out. A series of comprehensive and persuasive comparisons and experiments with different feature conversion methods (PCA and Isomap), different distance analysis algorithms (Euclidean distance and Mahalanobis distance) and different assessment models (HMM and DBN) verified that the proposed FKPC technique can monitor the operating status of the machine more accurately without complex model training and parameter selection processes.

REFERENCES

1. Joshua A, Sugumaran V. A data driven approach for condition monitoring of wind turbine blade using vibration signals through best-first tree algorithm and functional trees algorithm: a comparative study. ISA Trans. 2017;67:160-172.
2. Yampikulsakul N, Byon E, Huang S, Sheng S, You M. Condition monitoring of wind power system with nonparametric regression analysis. IEEE Trans Energy Convers. 2014;29(2):288-299.
3. He W, Miao Q, Azarian M, Pecht M. Health monitoring of cooling fan bearings based on wavelet filter. Mech Syst Signal Process. 2015;64:149-161.
4. Liu WY, Zhang WH, Han JG, Wang GF. A new wind turbine fault diagnosis method based on the local mean decomposition. Renewable Energy. 2012;48:411-415.
5. Huang Q, Jiang D, Hong L, Ding Y. Application of Wavelet Neural Networks on Vibration Fault Diagnosis for Wind Turbine Gearbox/International Symposium on Neural Networks. Berlin, Heidelberg: Springer; 2008:313-320.
6. Feng Z, Liang M. Complex signal analysis for wind turbine planetary gearbox fault diagnosis via iterative atomic decomposition thresholding. J Sound Vib. 2014;335(20):5196-5211.
7. Yuan N, Yang W, Kang B, Xu S, Li C, Signal fusion-based deep fast random forest method for machine health assessment. J Manuf Syst. 2018;48:1-8.
8. Cheng J, Zhang K, Yang Y. An order tracking technique for the gear fault diagnosis using local mean decomposition method. Mech Mach Theory. 2012;55:67-76.
9. Li C, Sanchez R-V, Zurita G, Cerrada M, Cabrera D, Va’squez RE. Multi-modal deep support vector classification with homologous features and its application to gearbox fault diagnosis. Neurocomputing. 2015;168:119-127.
10. Ottewill JR, Orkisz M. Condition monitoring of gearboxes using synchronously averaged electric motor signals. Mech Syst Signal Process. 2013;38(2):482-498.
11. Yu J. Health condition monitoring of machines based on hidden markov model and contribution analysis. IEEE Trans Instrum Meas. 2012;61(8):2200-2211.
12. Li C, Chen J, Hong R, Feng Y, Li Y. Degradation trend estimation of slewing bearing based on lssvm model. Mech Syst Signal Process. 2016;76:353-366.
13. Gan M, Wang C, Zhu. Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. Mech Syst Signal Process. 2016;72:92-104.
14. Li C, Liang M, Tang T. Criterion fusion for spectral segmentation and its application to optimal demodulation of bearing vibration signals. Mech Syst Signal Process. 2015;64:132-148.
15. Zhao M, Tang B, Tan Q. Bearing remaining useful life estimation based on time–frequency representation and supervised dimensionality reduction. Measurement. 2016;86:41-55.
16. Li C, Cabrera D, de Oliveira JV, Sanchez R-V, Cerrada M, Zurita G. Extracting repetitive transients for rotating machinery diagnosis using multiscale clustered grey infogram. Mech Syst Signal Process. 2016;76:157-173.
17. Hamel M, Addali A, Mba D. Investigation of the influence of oil film thickness on helical gear defect detection using acoustic emission. Appl Acoust. 2014;79:42-46.
18. Fan M, Zhang X, Qiao H, Zhang B. Efficient isometric manifold learning based on the self-organizing method. Inf Sci. 2016;345:325-339.
19. Hastie T, Stuetzle W. Principal curves. J Am Stat Assoc. 1989;84(406):502-516.
20. Gao Y, Wang M, Ji R, Wu X, Dai Q. 3-D Object retrieval with Hausdorff distance learning. IEEE Trans Industr Electron. 2013;61(4):2088-2098.

21. Pan Y, Chen J. The changes of complexity in the performance degradation process of rolling element bearing. J Vib Control. 2016;22(2):344-357.

22. Wang Y, Xu G, Liang L, Jiang K. Detection of weak transient signals based on wavelet packet transform and manifold learning for rolling element bearing fault diagnosis. Mech Syst Signal Process. 2015;54:259-276.

23. Jiang H, Chen J, Dong G, Wang R. An intelligent performance degradation assessment method for bearings. J Vib Control. 2016;23(18):1077546315624996.

24. Campbell K, Ponting CP, Webber C. Laplacian Eigenmaps and principal curves for high resolution pseudotemporal ordering of single-cell RNA-seq profiles. bioRxiv. 2015;027219.

25. Zhang Z, Wang J, Zha H. Adaptive manifold learning. IEEE Trans Pattern Anal Mach Intell. 2012;34(2):253-265.

26. Ghassabeh YA, Rudzicz F. Incremental algorithm for finding principal curves. IET Signal Proc. 2015;9(7):521-528.

27. Ma Y, Fu Y. Manifold Learning Theory and Applications. London, UK: CRC press; 2011.

28. Verbeek JJ, Vlassis N, Krose B. A k-segments algorithm for finding principal curves. Pattern Recogn Lett. 2002;23(8):1009-1017.

29. Thenmozhi S, Sumathi N. Clustering sentence level text using a hierarchical fuzzy relational and k-means clustering algorithm. Pattern Recogn. 2014;28(5):695-709.

30. Blömer J, Brauer S, Bujna K. Complexity and approximation of the fuzzy k-means problem. arXiv preprint arXiv: 1512.05947. 2015.

31. Zhu J, Huang Z, Peng X. Curve similarity judgment based on the discrete Fre'chet distance. J Wuhan Univ (Natural Science Edition). 2009;2:19-26.

32. Shen Y. On the probabilistic Hausdorff distance and a class of probabilistic decomposable measures. Inf Sci. 2014;263(1):126-140.

33. Zhao C, Shi W, Deng Y. A new Hausdorff distance for image matching. Pattern Recogn Lett. 2005;26(5):581-586.

How to cite this article: Yuan N, Yang W, Kang B, Xu S, Wang X, Li C. Manifold learning-based fuzzy k-principal curve similarity evaluation for wind turbine condition monitoring. Energy Sci Eng. 2018;6:727-738. https://doi.org/10.1002/ese3.247

Retracted