This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2018.2866549, IEEE Transactions on Industrial Informatics

Machine Health Monitoring Using Adaptive Kernel Spectral Clustering and Deep Long Short-Term Memory Recurrent Neural Networks

Yiwei Cheng, Haiping Zhu, Jun Wu, and Xinyu Shao

***Abstract*—Machine health monitoring is of great importance in industrial informatics field. Recently, deep learning methods applied to machine health monitoring have been proven effective. However, the existing methods face enormous difficulties in extracting heterogeneous features indicating the variation until failure and revealing the inherent high-dimensional features of massive signals, which affect the accuracy and efficiency of machine health monitoring. In this paper, a novel data-driven machine health monitoring method is proposed using adaptive kernel spectral clustering (AKSC) and deep long short-term memory recurrent neural networks (LSTM-RNN). This method include three steps: First, features in time domain, frequency domain and time-frequency domain are respectively extracted from massive measured signals. And, an Euclidean distance based algorithm is designed to select degradation features. Second, AKSC algorithm is introduced to adaptively identify machine anomaly behaviors from multiple degradation features. Third, a new deep learning model (LSTM-RNN) is constructed to update and predict the failure time of the machine. The effectiveness of the proposed method is validated using a set of test-to-failure experimental data. The results show that the performance of the proposed method is competitive with other existing methods.**

***Index Terms*—Machine health monitoring, deep long short-term memory recurrent neural networks (LSTM-RNN), adaptive kernel spectral clustering, anomaly detection, failure prognostics.**

I. INTRODUCTION

HEALTH monitoring of a machine and its components is crucial to avoid sudden failures, minimize machine downtime and unnecessary costs. Machine health monitoring is of great importance in industrial informatics field [1]. It uses measured signals obtained by built-in sensors to detect anomaly behavior and to predict failure time [2]. With rapid advances in sensors and communications technology, massive measured signals can be conveniently collected and transferred without a break [3], [4]. Thus, data-driven methods for machine health monitoring have attracted more and more attention in recent years.

Many data-driven methods like artificial neural networks [5], support vector machines [6], Bayesian networks [7], hidden Markov models [8], exponential regression [9], and stochastic processes [10] have been applied to machine health monitoring. However, these methods usually have weak adaptive learning capabilities and need spare much time to get optimal model parameters.

Deep learning, as a novel data-driven method, utilizes supervised/unsupervised machine learning mechanism to adaptive capture representative features from big data for the tasks of classification and pattern recognition [11]-[13]. It has been successfully applied to computer vision, natural language processing and speech recognition, causing the new round of high tide in machine learning field.

Some deep learning methods have already tried to find out corresponding solutions to machine health monitoring. In [14], a deep neural network (DNN)-based framework was developed to monitor conditions of wind turbine gearboxes and identify their impending failures. In [15], a multi-layer neural network with multi-valued neurons approach is utilized for deal with the reliability and degradation time series prediction problem. In [16], an integrated deep learning approach for multi-bearing remaining useful life collaborative prediction by combining both time domain and frequency domain features is proposed. In [17], a deep convolution neural networks (DCNN) is constructed for remaining useful life estimation by using the raw collected data. However, it is not easy for the deep learning methods such as DNN and CNN to extract heterogeneous features indicating the variation until failure of a machine since the massive measured signals are of complex multimodality and high heterogeneity. In addition, deep learning methods such as DNN and CNN run in vector space. Thus, it is very difficult to reveal the inherent high-dimensional features of the measured signals.

Manuscript received Sep. 29, 2017; revised Jan. 29, 2018, May 28, 2018 and Jun. 2, 2018; accepted Aug. 16, 2018. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant No. 51475189 and 51721092, in part by the Foundation of the National Key Intergovernmental Special Project Development Plan of China under Grant No. 2016YFE0121700, and in part by the Fundamental Research Funds for the Central Universities under Grant No. 2016YXMS050.(*Corresponding author: Jun Wu*.)

J. Wu is with School of Naval Architecture and Ocean Engineering, Huazhong University of Science and Technology and with Collaborative Innovation Center for Advanced Ship and Deep-Sea Exploration, Wuhan 430074, China (e-mail: wuj@hust.edu.cn).

Y.W. Cheng, H.P. Zhu, and X.Y. Shao are with National Engineering Research Center of Digital Manufacturing Equipment, Huazhong University of Science and Technology, Wuhan 430074, China (e-mails: chengyiwei102@163.com; haipzhu@hust.edu.cn; shaoxy@hust.edu.cn).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2018.2866549, IEEE Transactions on Industrial Informatics

Compared with other deep learning methods, long short-term memory recurrent neural network (LSTM-RNN) can pass machine state information across time steps allowing previously processed data to affect subsequent data to capture long-term dependencies and model sequential data, which effectively overcome the problems above. Some studies have been tried using LSTM for machine health monitoring. a LSTM neural network was proposed in [18] for RUL prediction of bearings by fusing multiple features to construct RNN based health indicator. In [19], LSTM-RNN was constructed for RUL estimation of aero-engines in the cases of complicated operations, hybrid faults and strong noises. Unlike the existing models, a deep learning based machine health monitoring method is proposed in this paper using adaptive kernel spectral clustering (AKSC) and long short-term memory recurrent neural network (LSTM-RNN) to identify incipient failure of machine and automatically determine the optimal time lags for machine failure prediction. Herein, AKSC algorithm is adopted to automatically and adaptively perceive the early failure of the machine. LSTM-RNN model is employed to capture long-term dependencies and model time series data through the deep learning network to realize failure prognostics from the beginning of anomaly state.

The main contributions of this paper are summarized as the following two points. First, adaptive kernel spectral clustering (AKSC) based anomaly detection method is proposed to perceive machine anomaly behaviors. A three-stage mechanism is devised to iteratively adjust the AKSC clustering model and adaptively identify machine incipient failure from multiple degradation features in time domain, frequency domain and time-frequency domain. Second, a new deep learning model is constructed by using LSTM-RNN to capture long-term temporal dependency for short-term failure time prediction, where multiple feature fusion is achieved with enhanced the accuracy and reliability of machine failure prognostics.

The rest of this paper is organized as follows. In Section II, feature extraction and selection are presented in detail. Section III is dedicated to the proposed data-driven machine health monitoring method. In Section IV, a practical experimental study is performed to demonstrate the effectiveness of the proposed method. Conclusions and future work of the research are discussed in Section V.

II. Background

The original sensor signals generally contain redundant information with noise and cannot be directly used to implement machine health monitoring. This section presents an overview of the feature extraction and selection procedure, with a focus on time-domain feature extraction, frequency-domain feature extraction, time-frequency feature extraction, and degradation feature selection.

A. Time-Domain Feature Extraction

Time-domain features have been proven effective for machine health monitoring [20]. In this paper, a total of 12 time-domain features are extracted from the waveforms of measured signals. Table I summarizes these time-domain features, where $x(t)$ is t-th sampling time, N is the number of the sampling times. These extracted features cover a wide range of popular time domain characteristics, with the features (tf_1, tf_2, tf_3, tf_4) reflecting the amplitude and energy over the time domain of a measured signals while the features (tf_5, tf_6, tf_7, tf_8, tf_9, tf_{10}, tf_{11}, tf_{12}) reflecting the distribution situation over the time domain.

TABLE I
TIME-DOMAIN FEATURES

No.	Formula	No.	Formula
1	$tf_1 = \frac{1}{N}\sum_{t=1}^{N}\lvert x(t)\rvert$	7	$tf_7 = \frac{tf_2}{\lvert\bar{x}\rvert}$
2	$tf_2 = \sqrt{\frac{1}{N}\sum_{t=1}^{N}x(t)^2}$	8	$tf_8 = \frac{tf_3}{tf_2}$
3	$tf_3 = \max(x(t))$	9	$tf_9 = \frac{tf_3}{\lvert\bar{x}\rvert}$
4	$tf_4 = \left[\frac{1}{N}\sum_{t=1}^{N}\sqrt{\lvert x(t)\rvert}\right]^2$	10	$tf_{10} = \frac{tf_3}{tf_4}$
5	$tf_5 = \frac{1}{N}\sum_{t=1}^{N}(x(t)-\bar{x})^3$	11	$tf_{11} = \frac{tf_6}{tf_2^4}$
6	$tf_6 = \frac{1}{N}\sum_{t=1}^{N}(x(t)-\bar{x})^4$	12	$tf_{12} = \frac{tf_5}{tf_2^4}$

B. Frequency-Domain Features Extraction

A fast Fourier transform is the most widely applied method for fault diagnostics in frequency domain [21], [22]. In this paper, measured signals are converted by fast Fourier transform into frequency spectrum. A total of 12 frequency-domain features are extracted from the frequency spectrum of measured signals. Table II summarizes these frequency-domain features, where $s(i)$ is the frequency spectrum of i-th sampling time, N is the number of spectrum lines, and f_i is the frequency value of i-th line. The features (ff_1, ff_2, ff_3, ff_4, ff_5) describe the variation of the main frequency band position over frequency domain of a measured signal, while the features (ff_6, ff_7, ff_8, ff_9, ff_{10}, ff_{11}, ff_{12}) describe the dispersion level of spectral energy distribution over frequency domain.

TABLE II
FREQUENCY-DOMAIN FEATURES

No.	Formula	No.	Formula
1	$ff_1 = \frac{\sum_{i=1}^{N}s(i)}{N}$	7	$ff_7 = \sqrt{\frac{\sum_{i=1}^{N}f_i^2 s(i)}{\sum_{i=1}^{N}s(i)}}$
2	$ff_2 = \frac{\sum_{i=1}^{N}(s(i)-ff_1)^2}{N-1}$	8	$ff_8 = \sqrt{\frac{\sum_{i=1}^{N}f_i^4 s(i)}{\sum_{i=1}^{N}f_i^2 s(i)}}$
3	$ff_3 = \frac{\sum_{i=1}^{N}(s(i)-ff_1)^3}{N(\sqrt{f_2})^3}$	9	$ff_9 = \frac{\sum_{i=1}^{N}f_i^2 s(i)}{\sqrt{\sum_{i=1}^{N}s(i)\sum_{i=1}^{N}f_i^4 s(i)}}$
4	$ff_4 = \frac{\sum_{i=1}^{N}(s(i)-ff_1)^4}{Nf_2^2}$	10	$ff_{10} = \frac{ff_6}{ff_5}$
5	$ff_5 = \frac{\sum_{i=1}^{N}f_i s(i)}{\sum_{i=1}^{N}s(i)}$	11	$ff_{11} = \frac{\sum_{i=1}^{N}(f_i-ff_5)^4 s(i)}{Nff_6^4}$
6	$ff_6 = \sqrt{\frac{\sum_{i=1}^{N}(f_i-p_5)^2 s(i)}{N}}$	12	$ff_{12} = \frac{\sum_{i=1}^{N}(f_i-ff_5)^{\frac{1}{2}}s(i)}{Nff_6}$

C. Time-Frequency Feature Extraction

Complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) is the latest version of empirical mode decomposition algorithm for signal processing in time-frequency domain. It can self-adaptively decompose a measured signal mixed with significant noises into a set of intrinsic mode functions (IMFs) and a residue [23], [24]. The frequency bands of these IMFs range from high to low, and the frequency components contained in each frequency band are different and change with the variation of the measured signal. When machine damages occur, the energy in the IMFs called

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2018.2866549, IEEE Transactions on Industrial Informatics

intrinsic mode energy will change accordingly. The intrinsic mode energy of machine damages is larger than normal one so that intrinsic mode energies as time-frequency features can be extracted using CEEMDAN. Given a preprocessed measured signal $x(t)$, the procedure of the intrinsic mode energies extraction is presented as follows:

First of all, a white noise series $w_i(t)$ is added to the signal $x(t)$. A new noise-added signal $x_i(t)$ is constructed as:

$$x_i(t) = x(t) + w_i(t),\ i = 1, 2,\ \dots, K \tag{1}$$

Then, all the local maximum and local minimum points of the signal $x_i(t)$ are checked and interpolated by cubic spline curves to form the upper envelope $y_{up}(t)$ and the lower envelope $y_{low}(t)$. The mean of the upper and lower envelopes is calculated as:

$$m_i(t) = \left(y_{up}(t) + y_{low}(t)\right)/2 \tag{2}$$

And, the mean $m_i(t)$ is subtracted by $x_i(t)$ to obtain a new sequence $h_i(t)$ for removing the low frequency component. If $h_i(t)$ satisfies the IMF condition, $h_i(t)$ is regarded as the first IMF component of $x_i(t)$ denoted $E_1\left(x_i(t)\right)$. Otherwise, $h_i(t)$ is assigned to $x_i(t)$ and the steps above need be repeated. Until now, the first IMF component of the K signals can be obtained. The mean of the first IMF components is computed as:

$$\widehat{IMF_1}(t) = \frac{1}{K}\sum_{i=1}^{K} E_1\left(x_i(t)\right) \tag{3}$$

Meanwhile, the residue $r_1(t)$ at the first stage is deduced as:

$$r_1(t) = x(t) - \widehat{IMF_1}(t) \tag{4}$$

If the residue $r_1(t)$ is monotonic, the decomposition will end. Otherwise, the residue $r_1(t)$ is added with different white noise $w_i(t)$ to reconstruct a new signal $x_{1i}(t)$, giving:

$$x_{1i}(t) = r_1(t) + \varepsilon_1 E_1\left(w_i(t)\right) \tag{5}$$

where ε_1 is signal-to-noise ratio. Following the previous steps, the signal $x_{1i}(t)$ is decomposed to obtain the second IMF component $\widehat{IMF_2}(t)$, which is described as:

$$\widehat{IMF_2}(t) = \frac{1}{K}\sum_{i=1}^{K} E_1\left(r_1(t) + \varepsilon_1 E_1\left(W_i(t)\right)\right) \tag{6}$$

Similarly, calculating all the IMF components and the final residue $r(t)$ of the signal $x(t)$, it can be reconstructed as:

$$x(t) = \sum_{\mathrm{j}=1}^{J} \widehat{IMF_{\mathrm{j}}}(t) + r(t) \tag{7}$$

where J is the number of the IMF components. The final residue $r(t)$ is regarded as the (J+1)-th IMF component $\widehat{IMF_{J+1}}(t)$.

Finally, the intrinsic mode energies of the machine are defined as:

$$IME_J = [1/(N-1)]\sum_{i=1}^{N}\left[\widehat{IMF_J}(t_i)\right]^2 \tag{8}$$

where N is the number of the sampling times.

D. Degradation Feature Selection

Since some extracted features above are not related to degradation phenomena of the machine, they cannot show the variation until failure. To avoid high dimensional problem and increase the effectiveness of machine health monitoring, degradation features need be picked up from extracted features. Euclidean distance is the most common use of measurement for distance [25]. In this paper, an Euclidean distance based algorithm is designed to select degradation features because that the distance between the same kind of degradation features is the smallest. The specific process is as follows:

First of all, these time-domain features, frequency-domain features and time-frequency features are merged into a set of feature vectors $\{\boldsymbol{F^{(1)}}, \boldsymbol{F^{(2)}}, \cdots, \boldsymbol{F^{(M)}}\}$. Each feature $\boldsymbol{F}^{(m)} = [\boldsymbol{x}_1^{(m)}, \boldsymbol{x}_2^{(m)}, \dots, \boldsymbol{x}_n^{(m)}]$ is n-dimensional observation vector. The features are treated as $\boldsymbol{M}$ points in n-dimensional space. The clustering center of the $\boldsymbol{M}$ points is calculated as:

$$\boldsymbol{F}^{\text{center}} = [\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \dots, \overline{\boldsymbol{x}}_n] = [\frac{1}{M}\sum_{m=1}^{M}\boldsymbol{x}_1^{(m)}, \frac{1}{M}\sum_{m=1}^{M}\boldsymbol{x}_2^{(m)}, \dots, \frac{1}{M}\sum_{m=1}^{M}\boldsymbol{x}_n^{(m)}] \tag{9}$$

Then, the Euclidean distance between m-th feature $\boldsymbol{F}^{(m)}$ and the clustering center $\boldsymbol{F}^{\text{center}}$ is computed as:

$$ED_m = \sqrt{\sum_{i=1}^{n}(\boldsymbol{x}_n^{(m)} - \overline{\boldsymbol{x}}_n)^2} \tag{10}$$

The smaller the Euclidean distance is, the more appropriate the feature is. The features whose Euclidean distance is less than a given threshold are filtered out.

III. Proposed Method

A. Framework

The framework of proposed data-driven machine health monitoring method with massive measured signals is presented in Fig. 1. The method includes three main steps as follows:

(1) Feature extraction and selection: This first step is aimed to extract and select features with degradation phenomena from measured signals. As mentioned in Section II, the extracted features include time-domain features, frequency-domain features, and time-frequency features. And, an Euclidean distance based algorithm is designed to filter out the most suitable degradation features.

(2) Anomaly detection: This second step is aimed to adaptively identify machine anomaly behavior from multiple degradation features in real time. An AKSC clustering model is constructed and three stages including initialization, calibration and detection are gradually implemented to iteratively adjust model parameters for matching future data evolution. The outlier indicator is defined to represent the degree to which the current state deviates from the normal state and measures the anomaly level.

(3) Failure prognostic: This third step is aimed to continuously update and predict failure time of the machine. A new LSTM-RNN model is built up to capture the long-term spatiotemporal dependency of multiple degradation features for short-term machine failure prognostics. Multiple feature fusion is achieved by the deep learning network with enhanced the accuracy and reliability of machine failure prognostics

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2018.2866549, IEEE Transactions on Industrial Informatics

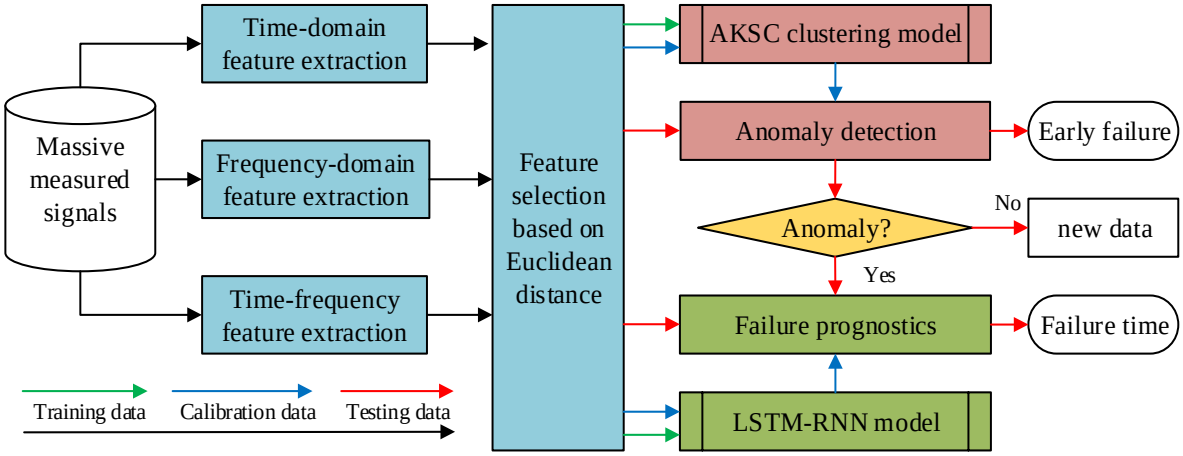

Fig. 1. Framework of machine health monitoring methodology.

B. Anomaly Detection based on AKSC Algorithm

An anomaly is defined to any event or signature that deviates from what is normal or expected. Machine damage might cause anomalous signals to emerge as a failure progresses. Early detection of an anomaly plays a vital role in machine heath monitoring. Due that anomalous signals in the early stage of machine damage are extremely weak and often mixed with much noise, accurate anomaly detection is a very big challenge. AKSC, as an adaptive and iterative spectral clustering method, can map original signals into a new feature space so as to find complex nonlinear cluster boundaries and identify the slight change in the features [26]-[31]. Thus, a novel anomaly detection method based on AKSC algorithm is proposed to identify incipient failure. The proposed method includes three stages: initialization, calibration and detection

1) Initialization Stage: In the first stage, an initial KSC clustering model is build up to map the training data into a new feature space of higher dimension, where groups of similar points can be perceived more easily.

Given a set of training data $\boldsymbol{F}_{tr} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N_{tr}}]$, where $\boldsymbol{x}_i \in \boldsymbol{R}^M$. The kernel function $\varphi(\cdot): \boldsymbol{R}^M \to \boldsymbol{R}^{M_h}$ is used to map the $\boldsymbol{F}_{tr}$ from the original dimension into the higher one. And the feature matrix $\boldsymbol{\Phi} = [\varphi(\boldsymbol{x}_1)^T; \ldots; \varphi(\boldsymbol{x}_{N_{tr}})^T]$ will be obtained. To get k clusters, a kernel spectral clustering algorithm is introduced to minimize the objective function as:

$$\min \frac{\sum_{l=1}^{k-1} w^{(l)^T} w^{(l)} - \sum_{l=1}^{k-1} \gamma^{(l)} e^{(l)^T} D^{-1} e^{(l)}}{2} \quad (11)$$

$$\text{Subject to} \begin{cases} e^{(1)} = \Phi w^{(1)} + 1_{N_{tr}} \otimes b_1 \\ \vdots \\ e^{(k\text{-}1)} = \Phi w^{(k\text{-}1)} + 1_{N_{tr}} \otimes b_{k\text{-}1} \end{cases}$$

where $w^{(l)} \in R^{M_h}$ represents the model parameter, b_l represents the bias term, $\otimes$ represents the Kronecker product, $e^{(l)} \in R^{N_{\text{tr}}}$ represents the projection of N_{tr} training data points in the space spanned by $w^{(1)}, w^{(2)}, \ldots, w^{(k\text{-}1)}$, D represents the degree matrix, and $\gamma^{(l)}$ represents regularization constant. In the process, an average membership strength criterion is used as performance function to select the parameters that maximize the separation between the clusters.

2) Calibration Stage: In the second stage, the parameters of the initial AKSC clustering model are automatically updated in order to match future data evolution and maximize the identification accuracy.

Given a set of calibration data $\boldsymbol{F}_{ca} = [x_1, \ldots, x_{N_{ca}}]$, where $\boldsymbol{x}_i \in \boldsymbol{R}^M$. The coordinates of each calibration data in eigenspace $c_{i,ca}$ are computed as:

$$\begin{cases} c_{i,ca} = \left\{ c_{i,ca}^{(l)} \right\}_{l=1}^{k-1} \\ c_{i,ca}^{(l)} = \dfrac{e_{i,ca}^{(l)}}{\gamma^{(l)} \sum_{j=1}^{N_{tr}} \alpha^{(l)} K(x_j, x_{i,ca})} \end{cases} \quad (12)$$

where $K(\cdot)$ is the radial basis function kernel function and $\alpha^{(l)}$ is the weight coefficient.

Then, the cluster membership for $x_{i,ca}$ can be computed as measured by the Euclidean distance between cluster centers and the location of data points in feature space. The clustering relation of the data points is determined. When new data points come in, the cluster centers will be updated.

3) Detection Stage: In the third stage, an outlier indicator is defined to represent the anomaly level and identify anomaly behavior of the machine in real time.

Given a set of testing data $\boldsymbol{F}_{te} = [x_1, \ldots, x_{N_{te}}]$, where $\boldsymbol{x}_i \in \boldsymbol{R}^M$. The outlier indicator is calculated as:

$$\text{OI}\left(x_{i,te}, c_i^{(l)}\right) = \frac{\max K(x_{i,te}, C^s) + \max K_\alpha\left(c_i^{(l)}, c_\alpha^s\right)}{2} \quad (13)$$

where OI represents the maximum similarity value between the data point $x_{i,te}$ and the cluster centers $c_i^{(l)}$. C^s represents the cluster centers in the eigenspace, and $K_\alpha(\cdot)$ represents the cosine similarity function in terms of the eigenvectors of the kernel matrix.

Generally, when the outlier indicator is lower than a given tolerance threshold, it means that current measurement data have a large difference with the normal state, in other words, machine anomaly appears.

C. LSTM-RNN Model for Failure Prognostics

To illustrate the variation until failure of a machine and its failure time, LSTM-RNN model will be constructed at the time that the machine anomaly behavior is identified. As is known, LSTM-RNN is a special kind of deep learning models where connections between neurons make a directed cycle. It is very suitable for processing data with sequential and time dependencies on multiple scales since the connections pass state information across time steps allowing previously

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2018.2866549, IEEE Transactions on Industrial Informatics

processed data to affect subsequent data. Thus, a new LSTM-RNN model is built to implement failure prognostics from the beginning of an anomaly detected in the previous section.

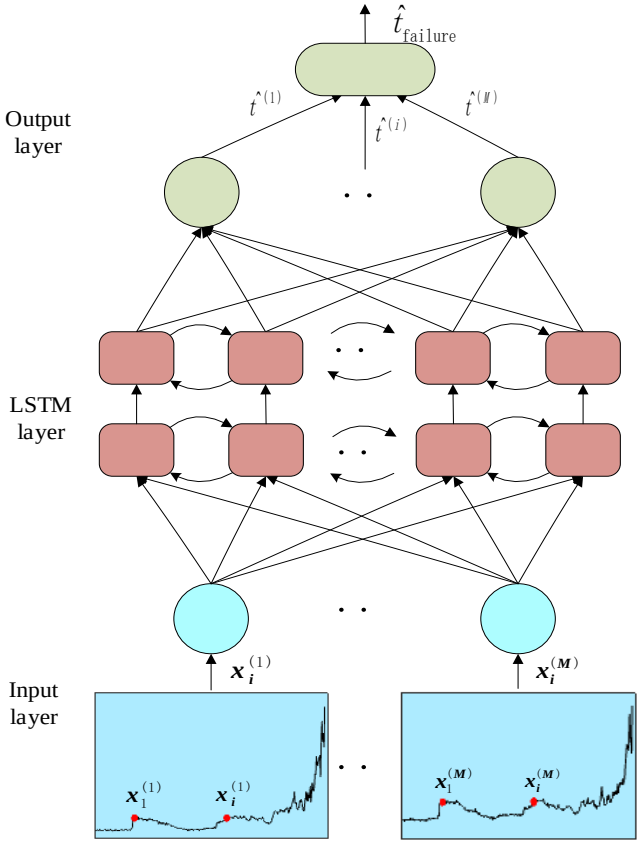

Fig. 2. Architecture of LSTM-RNN model for machine failure prognostics

As shown in Fig. 2, the LSTM-RNN model learns time series data of multiple features with long time spans and automatically determines the optimal time lags, which consists of one input layer, two LSTM layers and one output layer.

1) Input Layer: In the input layer, the number of input neurons is the same as that of the multiple features. The data set of the multiple features from the beginning of an anomaly are divided into training data, calibration data and testing data. The training data and calibration data are used to learn the LSTM-RNN model by back propagation through time algorithm, and the testing data are used to predict the failure time of the machine.

2) LSTM Layer: In the LSTM layer, memory cells connecting adjacent neurons form cycles that are self-connections of a neuron to itself across time. The input of memory cells includes the data x_i from the previous layer as well as the data $h_{i-1}$from themselves of the previous position. Accordingly, the output of the LSTM layer is affected not only by the current input data but also by the preceding state information.

The structure of a memory cell in the LSTM layer is shown in Fig. 3. The LSTM layer provide a fine nonlinear control mechanism over what is put into the memory and removed from the memory. Each memory cell has three gates to protect and control its state, including input gate, forget gate and output gate. The input gate determines which the cell state need be updated. The forget gate decides what information should be overlooked. The output gate resolves which part of the cell state will be exported. The specific formula of the memory cell are expressed as:

$$\begin{cases} g_i = \varphi\left(w_{gx}x_i + w_{gh}h_{i-1} + b_g\right) \\ n_i = \sigma(w_{nx}x_i + w_{nh}h_{i-1} + b_n) \\ f_i = \sigma\left(w_{fx}x_i + w_{fh}h_{i-1} + b_f\right) \\ o_i = \sigma(w_{ox}x_i + w_{oh}h_{i-1} + b_o) \\ s_i = g_i \odot n_i + s_{i-1} \odot f_i \\ h_i = \varphi(s_i) \odot o_i \end{cases} \tag{14}$$

where w_{gx}, w_{nx}, w_{fx} and w_{ox} are weight coefficients of the input x_i of the memory cell, respectively. w_{gh}, w_{nh}, w_{fh} and w_{oh} are weight coefficients of the previous output h_{i-1} of the memory cell, respectively. b_g, b_n, b_f, and b_o are bias of the input node, the input gate, the forget gate and the output gate, respectively. s_i and s_{i-1} is cell state values at time i and i-1, respectively. $\odot$ is the pointwise multiplication, σ are sigmod function and φ are the tanh function. The gradients of the weight coefficients and biases are optimized by stochastic gradient descent method.

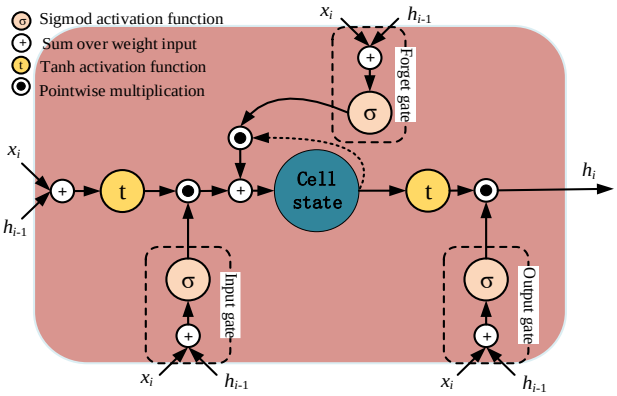

Fig. 3. Structure of a memory cell in the LSTM layer.

3) Output Layer: In the output layer, the failure time of the machine is deduced. For M multiple features, there are M failure time obtained, that is, $[\hat{t}^{(1)}, \hat{t}^{(2)}, \ldots, \hat{t}^{(M)}]$. We define the mean of these failure times as the predicted fault time, which is expressed as:

$$\hat{t}_{failure} = \frac{1}{M}\sum_{i=1}^{M} \hat{t}^{(i)} \tag{15}$$

IV. EXPERIMENTAL STUDY

A. Experimental Setup and Data Description

To validate effectiveness and reliability of the proposed method for machine health monitoring, vibration signals from test-to-failure experiments conducted by the NSF I/UCR Center for Intelligent Maintenance Systems are analyzed [32], [33]. The experiments are performed on a test rig as shown in Fig. 4. The test rig is composed of shaft, alternating current (AC) motor and rub belt. The shaft is driven by the AC motor via rub belts, with a slew speed of 2000 rpm and a radial load of 6000 lbs by a spring mechanism. Four thermocouples are installed on the outer ring of the tested bearings to measure the temperature of bearing lubricant oil. A magnetic sensor is mounted in the lubricant oil circuit. Once the amount of metal debris collected by the magnetic sensor exceeds a certain level, the experiment would stop.

All the three experiments are implemented on the test rig,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2018.2866549, IEEE Transactions on Industrial Informatics

which are denoted as Test 1, Test 2 and Test 3. In each experiment, four new bearings are installed on the shaft. Vibration signals are acquired by PCB 353B33 high sensitivity quartz acceleration sensor installed on the tested bearing and NI DAQ card 6062E data acquisition card. The sampling frequency is set to 20 kHz, and 20,480 points are recorded every 10 minutes.

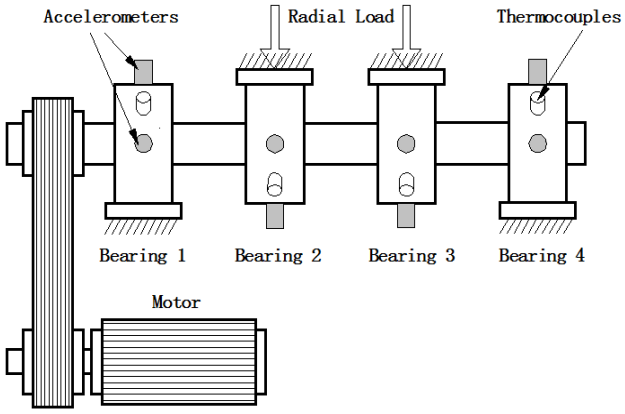

Fig. 4. Test rig and sensor placement illustration.

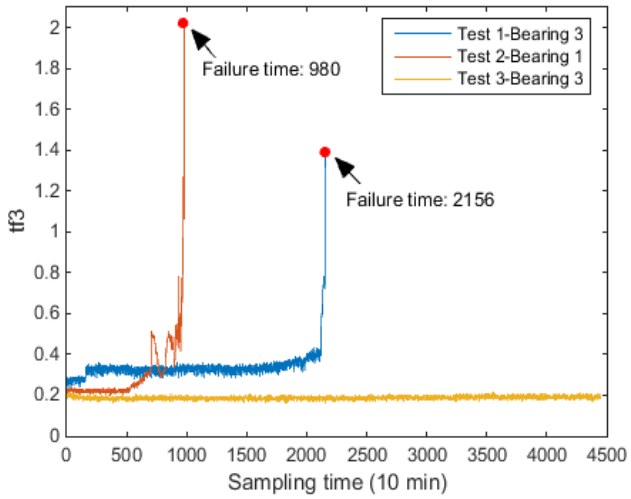

Fig. 5. Feature tf_3 extracted from three fault bearings.

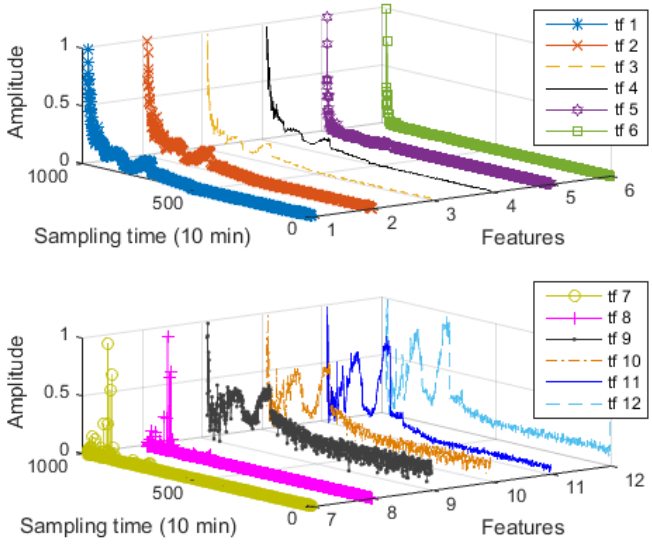

Fig. 6. Normalized time-domain features of the Test 2-Bearing 1.

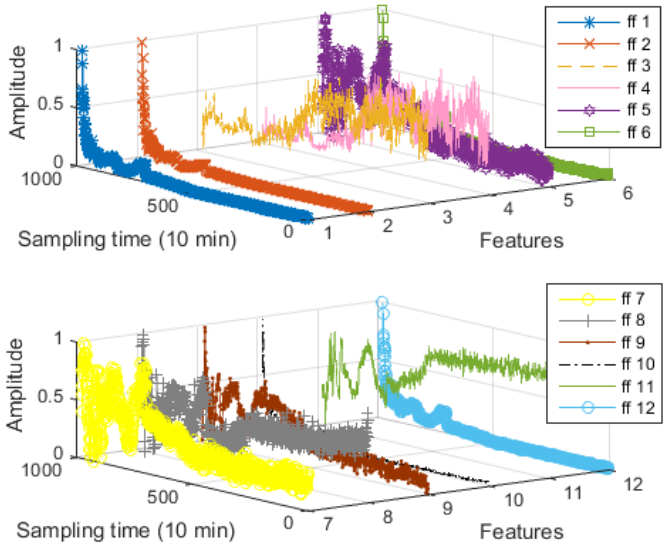

Fig. 7. Normalized frequency-domain features of the Test 2-Bearing 1.

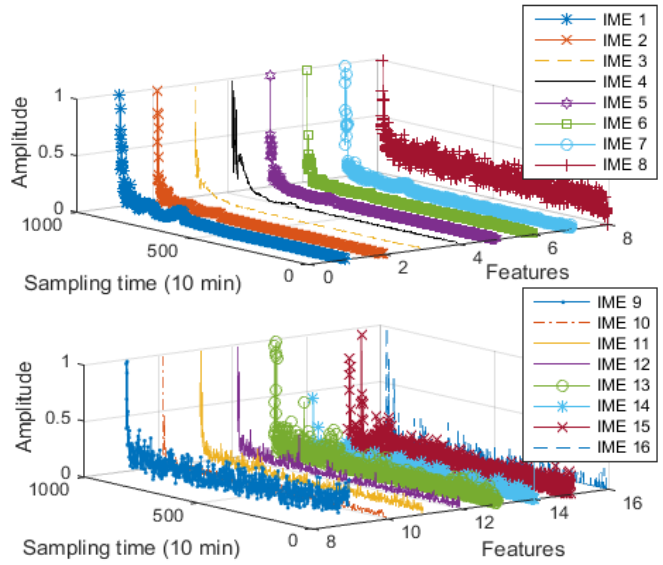

Fig. 8. Normalized time-frequency features of the Test 2-Bearing 1.

At the end of these experiments, inner race failure occurs in the bearing 3 of Test 1 (Test 1-Bearing 3), and outer race failure occurs in the bearing 1 of Test 2 (Test 2-Bearing 1) and the bearing 3 of Test 3 (Test 3-Bearing 3). The feature tf_3 in Table I is extracted from vibration signals of three fault bearings, which is shown in Fig. 5. It can be easily seen from Fig. 5 that Test 1-Bearing 3 and Test 2-Bearing 1 demonstrate different variation until failure; however, Test 3-Bearing 3 does not show such trend. So, the vibration signals from Test 1-Bearing 3 and Test 2-Bearing 1 are used in this paper.

In addition, all the proposed methods mentioned are implemented with MATLAB R2017a and executed on a computer with Intel Core i5-7500 (3.40 GHz) CPU, 16 GB memory and Microsoft Windows 10 enterprise operation system.

B. Feature Extraction and Selection

As described in Section II, time-domain features, frequency-domain features and time-frequency features are respectively

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2018.2866549, IEEE Transactions on Industrial Informatics

extracted from the vibration signals of Test 1-Bearing 3 and Test 2-Bearing 1.

Figs. 6-8 show the 40 extracted features of Test 2-Bearing 1. From Figs. 6-8, it can be observed that most of the features can well reflect the variation until failure, but some are not appropriate. Different features have different responses to failures. For example, the features ff_3 and ff_4 in Fig. 7 have strong volatility and do not describe degradation phenomena well.

Based on the proposed feature selection algorithm, the clustering center of 40 features is calculated, and the Euclidean distances between these features and the clustering center are obtained. The given threshold is set to 0.04. Fig. 9 shows the normalized Euclidean distance for Test 2-Bearing 1. From Fig. 9, it can be clearly seen that the features tf_3, ff_6, ff_{10}, ff_{12}, IME_8, IME_9, IME_{11} and IME_{12} are below the threshold level. Therefore, these 8 features were selected for the anomaly detection and the failure prognostics of Test 2-Bearing 1. In addition, the vibration signals from Test 1-bear 3 is processed in the same way, and the above 8 features are used for the anomaly detection and the failure prognostics of Test 1-bear 3.

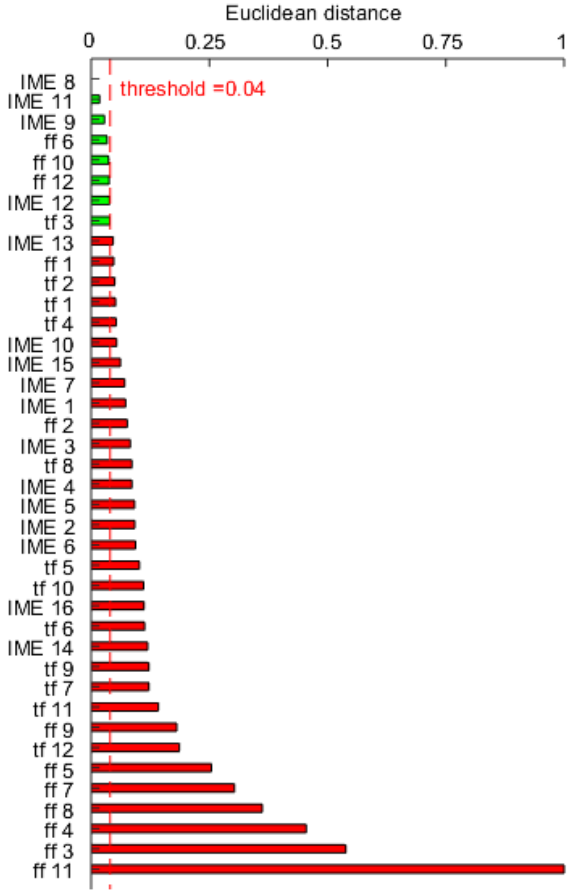

Fig. 9. Normalized Euclidean distance for Test 2-Bearing 1.

C. Anomaly Detection

According to the 8 selected features, AKSC-based anomaly detection method is used to perceive the anomaly behaviors of Test 2-Bearing 1 and Test 1-Bearing 3. The number of training data in the initialization stage and the number of calibration data in the calibration stage are both 250, and the remaining data are considered as the testing data and are inputted into the AKSC clustering model to identify anomaly behavior of the bearings. The given tolerance threshold for the anomaly detection is set to 0.0001. Fig. 10 shows the result of the anomaly detection of Test 2-Bearing 1. In Fig. 10(a), the clustering relationship between new measured data and the existing clusters is illustrated, and it can be seen from Fig. 10(a) that a new state is divided at the sampling time of 534. In Fig. 10(b), the OI value of every new measured data is shown, and it is obvious that the OI value of a new measured data at the sampling time of 534 is less than the threshold for the first time, which means that this measured data has very low similarity to all existing clustering centers. In Fig. 10(c), a representative of the measured data of the feature tf_3 is listed to verify the detection result, and it is easy to find that the value of the measured data after the sampling time of 534 has a larger increase than that of the previous sampling time, which means the anomaly appears at the sampling time of 534. In addition, it is quite similar to the result obtained by the cyclic energy indicator and the equivalent cyclic energy indicator in [34].

Fig. 11 shows the result of the anomaly detection of Test 1-Bearing 3. In Fig. 11(a), the OI value at the sampling time of 1810 is less than the threshold for the first time, this reflects a significant change in the data at the sampling time of 1810. The feature tf_3 of Test 1-Bearing 3 is used to confirm this result in Fig. 11(b), and the peak value of data fluctuation increases obviously after the sampling time of 1810.

TABLE III
CALCULATION TIME AND PEAK MEMORY OF PROPOSED METHOD

Bearing	Calculation time (s)	Peak memory (Kb)
Test 1-Bearing 3	6.901	6680
Test 2-Bearing 1	8.056	5228

In addition, calculation time and peak memory of the proposed anomaly detection method is shown in TABLE III. It can be observed from Table III that the proposed method can efficiently achieve the anomaly detection of the bearings on ordinary desktop computers.

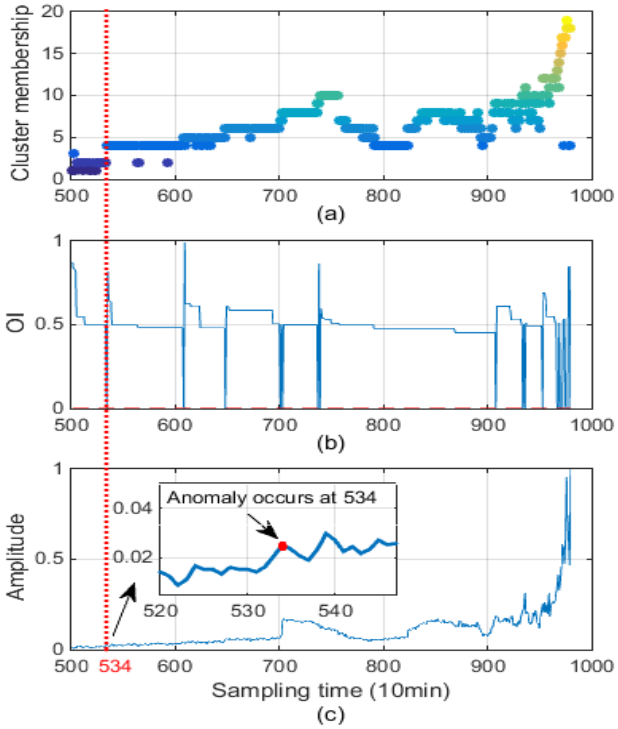

Fig. 10. Anomaly detection result of Test 2-Bearing 1: (a) the clustering relationship; (b) OI value; (c) tf3 feature.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2018.2866549, IEEE Transactions on Industrial Informatics

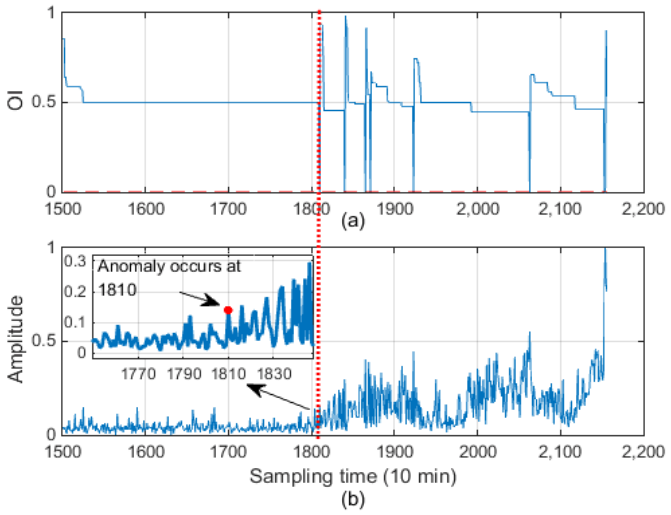

Fig. 11. Anomaly detection result of Test 1-bear 3: (a) OI value; (b) tf3 feature.

D. Failure Prognostics

Once the anomaly of the tested bearing is confirmed, the process of machine failure prognostics will be triggered based on the LSTM-RNN model.

In the training stage, a LSTM-RNN model is constructed with two LSTM layers and 17 LSTM memory cells in each layer. Each LSTM layer has a dropout with a value of 20%. The number of training epochs is set as 1000 and the batch size is set as 10. A mean square error function is chosen as loss function. After the LSTM-RNN model is set up, the 8 features obtained from Test 2-Bearing 1 are considered as a training set to train and validate the LSTM-RNN model. And the features after the sampling time of 534 are sequentially inputted into the LSTM-RNN model. Fig. 12 shows the predicted data and the true data of the training set. It can be seen from Fig. 12 that the predicted data have a high similarity with the true data, which reflects the effectiveness of the LSTM-RNN model. On this basis, the failure time of Test 2-Bearing 1 is deduced.

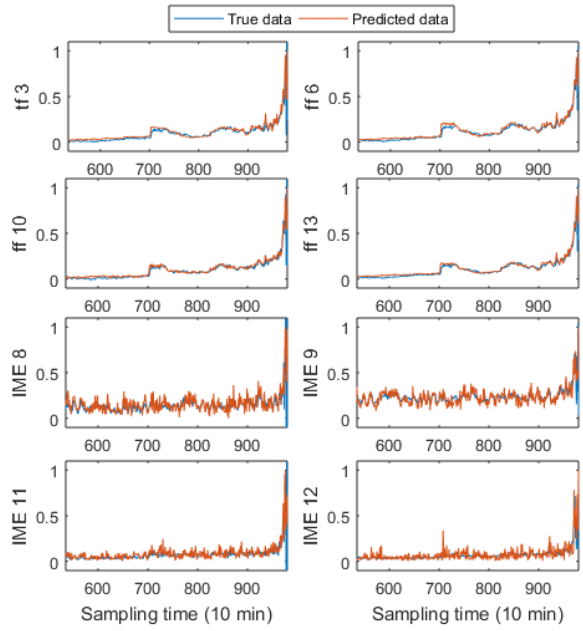

Fig. 12. Predicted data and the true data of 8 features of Test 2-Bearing 1.

TABLE IV
PREDICTED FAILURE TIME OF TWO BEARINGS BASED ON DIFFERENT FEATURE SELECTION METHODS.

Feature	Euclidean distance		Feature	Spearman's rank correlation coefficient	
	Test 2-Bearing 1	Test 1-Bearing 3		Test 2-Bearing 1	Test 1-Bearing 3
tf_3	979	2155	tf_3	980	2154
ff_6	979	2155	tf_5	980	2154
ff_{10}	979	2155	tf_6	980	2154
ff_{13}	979	2155	tf_{10}	977	2157
$IEMF_8$	977	2153	ff_2	980	2154
$IEMF_9$	979	2155	ff_5	981	2158
$IEMF_{11}$	979	2155	ff_7	973	2155
$IEMF_{12}$	984	2160	$IEMF_4$	980	2154
Average	979.375	2155.375	Average	978.875	2155
AC (%)	99.94	99.97	AC (%)	99.89	99.95

In the testing stage, the 8 features obtained from Test 1-Bearing 3 is used to test the trained LSTM-RNN model. The features after the sampling time of 1810 are sequentially inputted into the trained LSTM-RNN model and the failure time of Test 1-Bearing 3 is deduced. The final prediction results of Test 2-Bearing 1 and Test 1-Bearing 3 are shown in Table IV. It can be observed from Table IV that the accuracy of Test 2-Bearing 1 and Test 1-Bearing 3 are respectively 99.94% and 99.97%, which meet the demand in industrial applications.

In addition, the Euclidean distance based algorithm for degradation feature selection is compared with Spearman's rank correlation coefficient based method. They use the same model for training and testing. The results are shown in Table IV. It can be observed from Table IV that the Euclidean distance based algorithm has a better prediction accuracy than the Spearman's rank correlation coefficient based method.

Furthermore, three indicators including average error (AE), root mean square error (RMSE) and accuracy (AC) are applied to evaluate the performance of the methods. These indicators are expressed as:

$$\mathrm{AE} = \frac{1}{P}\sum_{i=1}^{P}|t(i) - t_r(i)| \tag{17}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{P}\sum_{i=1}^{P}(t(i) - t_r(i))^2} \tag{18}$$

$$\mathrm{AC} = \left(1 - \frac{|t_r - \hat{t}|}{t_r}\right) \times 100\% \tag{19}$$

where P is the number of prediction points, t_r is the true failure time, $t(i)$ is the predicted failure time, and $\hat{t}$ is the mean of the predicted failure times.

The proposed method is compared with other existing methods. Comparison of the prediction result of Test 1-Bearing 3 between LSTM-RNN and other existing methods are shown in Table V. It can be observed from Table V that the LSTM-RNN achieves the best prediction accuracy and lower error, which indicates its favorable generalization ability.

TABLE V
COMPARISON OF PROPOSED METHOD AND OTHER EXITING METHODS

Methods	AE	RMSE	AC (%)
Proposed method	6.03	6.07	99.97
Boukra [35]	N/A	N/A	93.44
Caesarendra et al. [36]	N/A	N/A	99.02
Widodo et al. [36]	N/A	N/A	98.41
Wu et al. [38]	N/A	N/A	99.42
Caesarendra et al. [39]	N/A	N/A	99.49
Qian et al. [40]	7.09	9.12	N/A

To analyze the complexity and overhead of the proposed method, the calculation time and peak memory are respectively counted, that is, 14.92s and 4096Kb. It indicates that the proposed method can efficiently realize the failure prognostics of the bearings on ordinary desktop computers.

V. Conclusions

In this paper, a novel data-driven machine health monitoring method is proposed using spectral clustering algorithm and deep learning model. The method aims to provide more effectiveness for perceiving machine anomaly behaviors and updating machine failure prediction. The framework of the proposed method is constructed with massive measured signals, which include three main steps: feature extraction and selection, anomaly detection, and failure prognostics. The first step targets on the extraction of multiple features in time domain, frequency domain and time-frequency domain, and the selection of features related to degradation phenomena of the machine. In the second step, an AKSC clustering model is established to identify machine incipient failure in real time. This is because AKSC algorithm can iteratively and adaptively adjust model parameters to future data evolution. In the third step, a new LSTM-RNN model is built up to continuously update and predict failure time of the machine starting from the identified incipient failure. This is due to its capability of processing data with sequential and time dependencies on multiple scales since the connections pass state information across time steps allowing previously processed data to affect subsequent data. The proposed method is validated using a set of test-to-failure experimental data. It is found that the proposed model provides high prediction accuracy due to its ability of capturing the long-term spatiotemporal dependency of multiple degradation features for short-term failure prognostics. In the comparative analysis, the proposed method exhibits great superiority over other existing methods in literature.

In addition, the machine health monitoring process generalizes as much as possible. The proposed method can be easily transferred to any other measured signals in industrial informatics field. However, due to time constraints and computational complexity, only the built deep architectures of the LSTM-RNN model is examined during this research. In the future, we will investigate different deep architectures of the LSTM-RNN model including variation in the depth of LSTM layer.

References

[1] S. Yin, X. Li, H. Gao, and O. Kaynak, "Data-based techniques focused on modern industry: An overview," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 657-667, Jan. 2015.

[2] M. H. Wei, M. Y. Chen, and D. H. Zhou, "Multi-sensor information based remaining useful life prediction with anticipated performance, " *IEEE Trans. Rel.*, vol. 62, no. 1, pp. 183-198, Mar. 2013.

[3] J. B. Yu, "Health condition monitoring of machines based on hidden Markov model and contribution analysis," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 8, pp. 2200-2211, Aug. 2012.

[4] A. Widodo, and B. S. Yang, "Machine health prognostics using survival probability and support vector machine," *Expert Syst. Appl.*, vol. 38, no. 7, pp.8430-8437, Jul. 2011.

[5] J. Wu, Y. H. Su, Y. W. Cheng, X. Y. Shao, C. Deng and C. Liu, "Multi-sensor information fusion for remaining useful life prediction of machining tools by adaptive network based fuzzy inference system," *Appl. Soft Comput.*, vol.68, pp. 12-23, Jul. 2018..

[6] V.T. Tran, H.T. Pham, B.S. Yang, T.T. Nguyen, "Machine performance degradation assessment and remaining useful life prediction using proportional hazard model and support vector machine," *Mech. Syst. Sig. Process.*, vol.32, pp.320-330. Oct. 2012.

[7] F. Q. Sun, N. Wang, X. Y. Li and W. Zhang, "Remaining useful life prediction for a machine with multiple dependent features based on bayesian dynamic linear model and copulas," *IEEE Access*, vol. 5, pp. 16277-16287, Aug. 2017.

[8] A. Soualhi, H. Razik, G. Clerc, D. D. Doan, "Prognosis of bearing failures using hidden markov models and the adaptive neuro-fuzzy inference system," *IEEE Trans. Ind. Electron.*, vol. 61, no. 6, pp.2864-2874, July 2013.

[9] J. Wu, C. Y. Wu, S. Cao, S. W. Or, C. Deng and X. Y. Shao, "Degradation data-driven time-to-failure prognostics approach for rolling element bearings in electrical machines," *IEEE Trans. Ind. Electron.*, DOI 10.1109/TIE.2018.2811366.

[10] P. Wang, B. D. Youn, and C. Hu, "A generic probabilistic framework for structural health prognostics and uncertainty management," *Mech. Syst. Signal Process.*, vol. 28, pp. 622-637, Apr. 2012.

[11] J. Deutsch, and D. He, "Using deep learning-based approach to predict remaining useful life of rotating components," *IEEE Trans. Syst., Man, Cybern., Syst.*, no. 99, pp. 1-10, May 2017.

[12] Z. Chen, and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Trans. Instrum. Meas.,* vol. 66, no. 7, pp. 1693-1702, Jul. 2017.

[13] L. Guo *et al.*, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomp*uting, vol. 240, pp. 98-109, May. 2017.

[14] L. Wang, Z. Zhang, H. Long, J. Xu and R. Liu, "Wind turbine gearbox failure identification with deep neural networks," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1360-1368, Jun. 2017.

[15] O. Fink, E. Zio and U. Weidmann, "Predicting component reliability and level of degradation with complex-valued neural networks," *Reliabil. Eng. Syst. Saf.,* vol. 121, no. 1, pp. 198-206, Jan. 2014.

[16] L. Ren, J. Cui, Y. Sun and X. Cheng, "Multi-bearing remaining useful life collaborative pre- diction: a deep learning approach," *J. Manuf. Syst.*, vol. 43, pp. 248-256, Mar. 2017.

[17] X. Li, Q. Ding and J. Q. Sun, "Remaining useful life estimation in prognostics using deep convolution neural networks," *Reliabil. Eng. Syst. Saf.,* vol. 172, pp. 1-11, Apr. 2018.

[18] L. Guo, N. P. Li, F. Jia, Y. G. Lei and J. Lin, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomputing*, vol. 240, pp.98-109, May 2017.

[19] Y. T. Wu, M. Yuan, S. P. Dong, L. Lin and Y. Q. Liu, "Remaining useful life estimation of engineered systems using vanilla LSTM neural networks," *Neurocomputing*, vol. 275, pp.167-179, Jan. 2018.

[20] J. Wu, C. Y. Wu, Y. Q. Lv, C. Deng, and X. Y. Shao, "Design a degradation condition monitoring system scheme for rolling bearing using EMD and PCA," *Ind. Manage. Data Syst.*, vol. 117, no. 4, pp.713-728, 2017.

[21] A. Ghods, and H.H. Lee, "Probabilistic frequency-domain discrete wavelet transform for better detection of bearing faults in induction motors," *Neurocomputing*, vol.188, pp.206-216, May 2016.

[22] R. Singleton, E. Strangas, and S. Aviyente, "The use of bearing currents and vibrations in lifetime estimation of bearings," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1301-1309, Jun. 2017.

[23] Y. H. Wang, C. Deng, J. Wu, Y. C. Wang, and Y. Xiong, "A corrective maintenance scheme for engineering equipment," *Eng. Fail. Anal.*, vol. 36, pp. 269-283, Jan. 2014.

[24] Y. B. Li, M. Q. Xu, X. H. Liang, and W. H. Huang, "Application of bandwidth EMD and adaptive multiscale morphology analysis for incipient fault diagnosis of rolling bearings," *IEEE Trans. Ind. Electron.*, vol. 64, no. 8, pp. 6506-6517, Jan. 2017.

[25] J. Tian, C. Morillo, M. H. Azarian, and M. Pecht, "Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with k-nearest neighbor distance analysis," *IEEE Trans. Ind. Electron.*, vol. 63, no. 3, pp. 1793-1803, Mar. 2016.

[26] R. Langone et al., "Automated structural health monitoring based on adaptive kernel spectral clustering," *Mech. Syst. Signal Process.*, vol. 90, pp. 64-78, Jun. 2017.,

[27] C. Alzate, J.A.K. Suykens, "Out-of-sample eigenvectors in kernel spectral clustering," in: Proc. of the International Joint Conference on Neural Networks (IJCNN 2011), 2011, pp. 2349-2356.

[28] C. Alzate, J. A. K. Suykens, "Multiway spectral clustering with out-of-

sample extensions through weighted kernel PCA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 335-347, Feb. 2010.

[29] A. Sarkheyli, A.M. Zain, and S. Sharif, "Robust optimization of ANFIS based on a new modified GA," *Neurocomputing* vol. 166, pp. 357-366, 2015.

[30] Y. Wang, Y. Peng, Y. Zi, X. Jin, and K. L. Tsui, "A two-stage data-driven-based prognostic approach for bearing degradation problem," *IEEE Trans. Ind. Informat.*, vol. 12, no. 3, pp. 924-932, Jun. 2016.

[31] Y. H. Wang, C. Deng, J. Wu, and Y. Xiong, "Failure time prediction for mechanical device based on the degradation sequence," *J. Intell. Manuf.*, vol. 26, no. 6, pp. 1181-1199, 2015.

[32] H. Qiu, J. Lee, J. Lin, and G. Yu, "Robust performance degradation assessment methods for enhanced rolling element bearing prognostics," *Adv. Eng. Inform.*, vol. 17, pp. 127-140, 2003.

[33] H. Qiu, J. Lee, J. Lin and G. Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *J. Sound Vibrat.*, vol. 289, no. 4, pp. 1066-1090, Feb. 2006.

[34] D. Wang, and C. Q. Shen, "An equivalent cyclic energy indicator for bearing performance degradation assessment," *J. Vib. Control*, vol. 22, no. 10, pp. 2380-2388, Jun. 2014.

[35] T Boukra, Identifying New Prognostic Features for Remaining Useful Life Prediction Using Particle Filtering and Neuro-Fuzzy System Predictor. In 2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC), Rome, Italy. Jun. 10-13, 2015.

[36] W. Caesarendra, A. Widodo, and B. S. Yang, "Combination of probability approach and support vector machine towards machine health prognostics," *Probabilist. Eng. Mech.*, vol. 26, no. 2, pp. 165-173, Apr. 2011.

[37] A. Widodo and B. S. Yang, "Machine health prognostics using survival probability and support vector machine," *Expert Syst. Appl.*, vol. 38, no.7, pp. 8430-8437, Jul. 2011.

[38] B. Wu, W. Li, M. Q. Qiu, "Remaining Useful Life Prediction of Bearing with Vibration Signals Based on a Novel Indicator," *Shock Vib.*, vol. 2017, pp. 1-10, Oct. 2017.

[39] W. Caesarendra, A. Widodo, and B. S. Yang, "Application of relevance vector machine and logistic regression for machine degradation assessment," *Mech. Syst. Signal Process.*, vol. 24, no. 4, pp. 1161-1171, May. 2010.

[40] Y. Qian, R. Yan, and R. X. Gao, "A multi-time scale approach to remaining useful life prediction in rolling bearing," *Mech. Syst. Signal Process.*, vol. 83, pp. 549-567, Jan. 2017.

Yiwei Cheng received his B.S. degree in Marine engineering from Dalian Maritime University, China, in 2016.

He is currently a Ph.D candidate at School of Mechanical Science and Engineering, Huazhong University of Science and Technology (HUST), China. His main research interests include big data analytics, health monitoring and remaining useful life prediction for equipment.

Haiping Zhu received his B.S. degree from Gansu University of Technology, China, in 1998, and received his M.S. and Ph.D degrees in mechanical engineering from Huazhong University of Science and Technology (HUST), in 2001 and 2005, respectively.

He is currently a Full Professor of School of Mechanical science and Engineering at HUST. His research interests include modeling and optimization of manufacturing system, reliability analysis and maintenance decisions, intelligent manufacturing and digital workshop applications. He has more than 80 publications and the award of 8 patents, and receives several awards for his teaching activities.

Jun Wu received his B.S. degree from Hubei University of Technology, China, in 2001, and received his M.S. and Ph.D. degrees in mechanical engineering from Huazhong University of Science and Technology (HUST), in 2004 and 2008, respectively.

He is currently an Associate Professor of School of Naval Architecture and Ocean Engineering at HUST. He worked as a visiting scholar at Stanford University, CA, USA from 2014 to 2015, where he conducted technical research in the area of structure health monitoring. His research interests include equipment health monitoring, fault diagnosis and remaining useful life prediction. He has more than 50 publications and the award of 9 patents, and receives several awards for his teaching activities.

Xinyu Shao received his B.S., Ph.D. degrees in mechanical engineering from Huazhong University of Science and Technology (HUST), in 1990 and 1998, respectively. He was a visiting Ph.D. candidate with the Department of Industrial and Manufacturing Systems Engineering, University of Michigan–Dearborn, Dearborn, MI, USA, from 1995 to 1998.

He is currently a Full Professor with the School of Mechanical Science and Engineering, HUST. He is also the Director of the National Engineering Research Center for the Digitization of Manufacturing Equipment. His research interests include digital and intelligent manufacturing, concurrent engineering, and quality/reliability engineering. He was a recipient of the Ministry of Education's Chang Jiang Scholars Program Professor in 2004, the National Science Fund for Distinguished Young Scholars in 2008, and the Second-Grade State Scientific and Technological Progress Prizes in 2001, 2008, and 2012.