International Journal of Parallel Programming (2020) 48:61–79
https://doi.org/10.1007/s10766-019-00650-1

Empirical Mode Decomposition and Temporal Convolutional Networks for Remaining Useful Life Estimation

Wensi Yang[1,4] · Qingfeng Yao[2,4] · Kejiang Ye[1] · Cheng-Zhong Xu[3]

Received: 10 August 2019 / Accepted: 4 November 2019 / Published online: 14 November 2019

Abstract

Remaining useful life (RUL) prediction plays an important role in guaranteeing safe operation and reducing maintenance cost in modern industry. In this paper, we present a novel deep learning method for RUL estimation based on time empirical mode decomposition (EMD) and temporal convolutional networks (TCN). The proposed framework can effectively reveal the non-stationary characteristics of bearing degradation signals and acquire time-series degradation signals which namely intrinsic mode functions through empirical mode decomposition. Furthermore, the feature information is used as the input to convolution layer and trained by TCN to predict remaining useful life. The proposed EMD–TCN model structure maintains a superior result compared to several state-of-the-art convolutional algorithms on public data sets. Experimental results show that the average score of EMD–TCN model is improved by 10–20% than traditional convolutional algorithms.

Keywords Convolutional neural networks · Empirical mode decomposition · Remaining useful life · Reliability

✉ Kejiang Ye
kj.ye@siat.ac.cn

Wensi Yang
ws.yang@siat.ac.cn

Qingfeng Yao
yaoqingfeng@sia.cn

Cheng-Zhong Xu
czxu@um.edu.mo

1 Shengzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shengzhen, China

2 Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China

3 Department of Computer and Information Science, University of Macau, Macau, China

4 University of Chinese Academy of Sciences, Beijing, China

1 Introduction

Modern industrial enterprises usually need to maintain their production equipment in well working condition for a long time to remain competitive. It is critical to improving the availability, stability, and safety of equipment under the premise of reducing equipment maintenance loss. Therefore, how to predict the equipment's useful life becomes a crucial task. Accurate equipment life prediction can provide equipment state warnings for maintenance operations in advance. In consequence, the unexpected downtime and enterprise losses are also reduced for the rational arrangement of maintenance personnel, which is also significant for attendant social and economic benefits. Currently, Condition-Based Maintenance (CBM) and Predictive Maintenance (PM) are the most effective methods [1]. Contrary to the traditional maintenance after the failure, these two methods optimize equipment maintenance strategies via diagnosis and prognosis of faults. In the CBM or PM, the maintenance of the equipment is based on observing or predicting the health status of equipment.

Typically, a CBM system includes seven parts: sensor, signal processing system, fault detection system, health assessment system, fault prediction system, test support system, and finally display part, where failure prediction is a core part. Recently, more and more attention of academy has been focused on failure prediction methods, as well as the industrial field [2]. And, numerous failure prediction methods have been proposed, which can be divided into two categories: *model-based methods* and *data-driven methods* [3]. The model-based prediction methods rely on analytical models to describe the operating state of industrial equipment [4]. However, the aging mechanism of real device systems is usually non-linear, randomized and dynamic, which causes difficulty to obtain accurate results by an analytical model [5]. The data-driven approach is designed to transform the device's detection and operational data into the degradation information of the device, which reveals the system operational status and corresponding degradation mechanism model [6]. Such methods exploit Artificial Intelligence (AI) and statistical methods to learn the degradation patterns of devices and predict the remaining useful life (RUL) of the device [7]. Data-driven methods can be applied to the scenario which is difficult to build analytical model [8], they are effective to transform noisy data into logical information for remaining useful life estimation. The proposed EMD–TCN framework in this paper is a kind of data-driven method.

Machine learning is the mainstream method in today's data-based prediction methods. Convolutional neural network (CNN), a model which can learn a high-level representation of data, is wildly used for the excellent application, such as the visual target recognition, machine translation, and prediction of RUL [9]. And one type variety of CNN is long short-term memory (LSTM), which is also used to predict useful life due to it can remain the recent memories of input. The condition of the commercial unit not only relate to current data but also relate to history data recently. Recently, with the rapid development of deep learning, a new architecture had been invented—temporal convolutional networks (TCN) which inspired by both memories of LSTM and have the ability to extract abstract features as CNN.

Empirical mode decomposition (EMD), is a signal decomposition method that can be effective for energy time-series forecasting [10]. EMD implements a sifting algorithm to adaptively decomposes time-series signals to its AM/FM modulated subcomponents. These subcomponents are called intrinsic mode functions (IMFs).

The contributions of this paper can be summarized as follows:

(1) Empirical mode decomposition algorithm was applied to extract more identifiable features of original signal for RUL estimation. The local features at different time scales are maintained to enhance the prediction accuracy of the network.
(2) The proposed systematic approach combines temporal convolutional network (TCN) method and empirical mode decomposition algorithm into a framework, which could realize a data-driven fault prediction model to estimate RUL.
(3) The proposed method is evaluated on Prognostics and Health Management 2012 (PHM2012) data [11]. Experimental results show that our framework makes good utilization of the inherent information of the data, and provides the RUL of the device in advance of failure.

The rest of the paper is organized as follows. In Sect. 2, related work about remaining useful time estimation is discussed. Section 3 presents the theoretical background of the proposed framework. Section 4 provides an analysis of the dataset and experimental results. Some concluding remarks and recommendations for future work are summarized in Sect. 5.

2 Related Work

The reliability, availability, and safety of a system are determining factors in the effectiveness of industrial performance [12]. Hence, predicting the remaining useful life (RUL) before the failure occurs, given the current machine condition to help engineers to reasonably judge the working state of equipment is crucial. The main methods for estimating the remaining useful life can be divided into a model-based approach and a data-driven approach.

The model-based approach assumes that an accurate physical degradation model can be obtained to predict the development of the failure process [13].A series of model-based methods have been proposed for RUL prediction. A comprehensive prognostic process based on data collected from model-based simulations under nominal and degraded conditions is described [14]. In this process, the remaining useful life prediction is obtained by mixing mode-based life predictions through time-averaged mode probabilities. Li *et al*. [15] present a stochastic defect propagation model for the remaining useful life prediction of defective bearings. A model-based diagnostic system [16] consists of a number of nonlinear models representing a set of rolling element bearing faults. However, in practical complex production systems, the aging mechanism of the system is usually random and difficult to obtain. Model-based methods are usually difficult to build accurate physical degradation models and the parameters of the model require extensive experiments and empirical data to determine.

The data-driven method is designed to transform the raw monitoring data into relevant information related to system degradation process and the degradation models are derived without concerning about the physics of the system degradation processes [17]. The data-driven approach usually consists of two phases: the first phase learns the fault degradation process and then the second phase predicts the future state of the fault [18]. The data-driven methods mainly use artificial intelligent (AI) approaches and statistical approaches to learn the degradation patterns and estimate the remaining useful life of devices. The statistical model-based approaches require the use of parameters of the engineered system under consideration to formulate a statistical model. The data gathered over a period of time are used as input to this model and the remaining useful life is determined. Banjevic *et al*. [19] discussed the RUL estimation for a Markov failure time process, including the joint model of the proportional hazard model (PHM) and the Markov property of the covariate evolution as a special case. Sankararaman et al. [20] investigated the use of the inverse first-order reliability method (inverse-form) to quantify the uncertainty in the remaining useful life. The inverse-form method is used to quickly obtain probability bounds of the remaining useful life estimation, and then the overall entire probability distribution of remaining useful life estimation is calculated.

Artificial intelligence technology has been increasingly applied to machine diagnosis and RUL estimation. A kind of popular AI techniques for RUL estimation is artificial neural networks. The goal of these methods is either to formalize the description of the transformation rules that link the inputs and outputs or to construct a prototype that exhibits behavior approximate to the system. Neural network as a kind of parallel computing model has good nonlinear mapping ability and are able to model extremely complex functions. They are easy to use, robust, and have the ability to generalize. Hence, they are effective to predict the remaining useful life. A TSNN model [21] for RUL prediction, performs an additive latent failure risk estimation and multiple binary classifications for predicting RUL-specific probabilities, which is optimized by minimizing the censoring KL divergence between the actual survival process and the resulting probabilities. Tian et al. [22] developed a robust estimation approach for RUL based on the ANN. The age and multiple condition monitoring measurement values at the present and previous inspection points as the inputs of the ANN model and the ANN model will estimate the life percentage as an output. In order to reduce the influence of the noise factors in the signal, the data is fitted using a generalized function of the Weibull failure rate function. The entire degradation process is taken into consideration by this RUL prediction approach. Since there are significant differences in the patterns of the changes at different degradation stages, it is difficult to fit the entire degradation process with a single accurate model. Soualhi et al. [23] presented an approach that combined Hilbert-Huang transform (HHT), the support vector machine (SVM), and the support vector regression (SVR) for the monitoring of ball bearings degradation. Remaining useful life was obtained by using a one-step time-series prediction based on SVR.

Our group at SIAT, CAS also proposed a lot of algorithms and tools for fault prediction and anomaly detection. An AMF–LSTM [24] algorithm for abnormal traffic detection in network by using deep learning model was presented. Zhu [25] presented an anomaly detection method based on feedforward neural network (FNN) model and convolutional neural network (CNN), which has a strong ability for network anomaly detection. A dynamic network anomaly detection system [26] by combining long short term memory (LSTM) and attention mechanism (AM) was designed to maintain the security of the network. Ye [27] proposed fault detection models for detecting the fault behaviors and interferences of up-to-date AI applications in container-based cloud systems. Firstly, the container-based cloud system fault injection framework was proposed. Then, based on the quantile regression method, the fault detection model was designed to detect potential faults in containers. While this paper focuses on RUL prediction.

3 Methodology

3.1 Empirical Mode Decomposition

In 1998, Huang et al. [10] proposed empirical mode decomposition (EMD) method to decompose time-series data without any limitation on the character, which has a significant superiority in handling non-stationary and nonlinear data. Therefore, the EMD method has been applied effectively in multiple engineering fields such as the ocean, atmosphere, celestial and geophysical data analysis.

The principle of the EMD method is to decompose time-series signals into a finite number of intrinsic mode functions (IMFs) which represent various components of signals containing local characters in different time scales. An IMF must satisfy two necessary criteria: one is that the number of local extreme points and zero-crossing points must be equal or at most one difference in the whole time scale. The other is that at any time point, the envelope mean value of local maximum value and local minimum value must be zero or close to zero.

The essence of the empirical mode decomposition method is to obtain the intrinsic fluctuation pattern by characterizing the time scale of the data and then decompose the data. The decomposition process is:

(a) For a given original signal data x(t), finding all the maxima points of the original data sequence x(t) and using the cubic spline interpolation function to fit the upper envelope $e_{max}(t)$ of the original data. Similarly, finding all the minimum values to fit the lower envelope $e_{min}(t)$.

(b) The mean of the upper envelope $e_{max}(t)$ and the lower envelope $e_{min}(t)$ is recorded as m_1:

$$m_1 = \frac{1}{2}(e_{max}(t) + e_{min}(t)) \tag{1}$$

(c) After subtracting the mean from the original signal data, a new data sequence $h_1(t)$ is obtained from x(t) by Eq. (2) :

$$h_1 = x(t) - m_1. \tag{2}$$

(d) If h_1 satisfies the oscillating mode condition [28], it is taken as the first IMF and recorded as $f_1(t) = h_1(t)$.
If h_1 does not satisfy the condition, repeat the above step (a) to (c) and regard $h_1(t)$ as the analysis signal until the IMF generation condition is met. The first IMF obtained is recorded as $f_1(t)$.

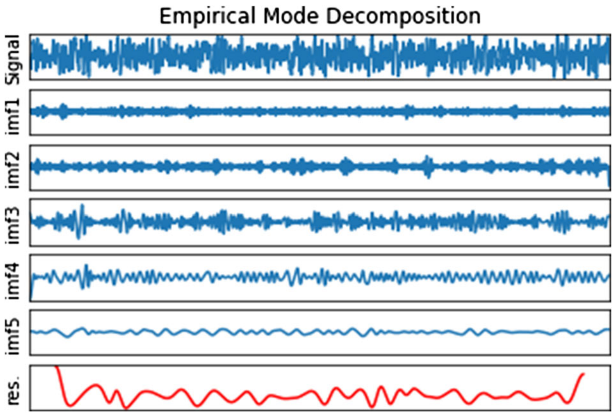


Fig. 1 An example of empirical mode decomposition of original signal

(e) Separate $f_1(t)$ from the original signal data and compute residue signal :

$$r_1(t) = x(t) - f_1(t) \tag{3}$$

the residue signal $r_1(t)$ as new original signal data, repeat steps (a) to (d) to separate a series of components that meet the IMF conditions :

$$r_2(t) = r_1(t) - f_2(t)$$
$$\cdot$$
$$\cdot$$
$$r_n(t) = r_{(n-1)}(t) - f_n(t)$$

When $r_n(t)$ cannot satisfy the IMF condition, the decomposition process terminates and the original signal data x(t) is decomposed as follows:

$$x(t) = \sum_{i=1}^{n} f_i(t) + r_n(t) \tag{4}$$

where n is the number of IMFs, $f_i(t)$ is the IMFs and $r_n(t)$ represents the residue. Figure 1 depicts an example of empirical mode decomposition of original signal data.

3.2 Temporal Convolutional Network

Temporal convolutional network (TCN) [29] is a kind of sequential prediction model that is designed to learn hidden temporal dependencies within input sequences. It is a simple convolutional architecture outperforms canonical recurrent neural networks. TCN takes the sequence $(x_0, x_1, \ldots, x_T)$ as the inputs, and the corresponding $(y_0, y_1, \ldots, y_T)$ as the expected outputs. Formally, a sequence produced by model such as function $f : X^{T+1} \rightarrow Y^{T+1}$ that mapping:

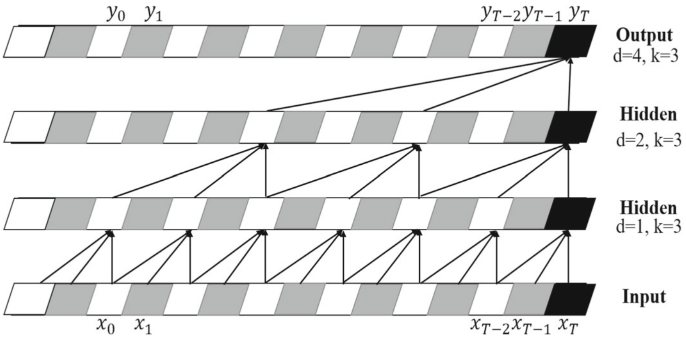


Fig. 2 Architecture elements of TCN, d is dilation factor and k is filter size

$$y_0, y_1, \ldots, y_T = f(x_0, x_1, \ldots, x_T)$$

TCN is a deep convolutional architecture characterized by layered stacks of dilated causal convolutional filters with residual connections [29]. Causal convolutions allow the model to make predictions on continuous streaming trace data, this character is necessary for Prognostics and Health Management (PHM). Dilated convolutions allow precise control over the receptive field while residual connections enable the model to have high-capacity and stable training. The architecture of TCN is as follows:

In general, multi-dimension sensor signals can be directly used as the input of the convolution layer. TCN can discover the intrinsic relationship between signal features with higher prediction accuracy and avoiding personal bias. It turns out that TCN has achieved good results and application in computer vision, such as image classification, target detection. In this paper, we exploit the potential of TCN in the fields of Prognostics and Health Management (PHM). A temporal convolutional layer consists of 1-D fully-convolutional network (FCN) and causal convolutions. 1-D fully-convolutional network (FCN) enables the length of each hidden layers as same as the length of the input layer. Causal convolutions, where output at time t can only be obtained from the convolution operation of $t-1$ and previous time step, so that they can ensure that the prediction of the time t does not use future information [30].

3.2.1 1-D Fully-Convolutional Network (FCN)

Let $x^{\ell-1}$ be the input of the ℓth layer and x^{ℓ} be the output of the ℓth layer, since one layer has multiple feature maps, we use x_j^{ℓ} represents the jth feature map of layer ℓ, and it can be obtained by following formula:

$$x_j^{\ell} = f\left(\sum_{i=1} x_j^{\ell-1} * w_{i,j}^{\ell} + b_j^{\ell}\right) \tag{5}$$

$w_{i,j}^{\ell}$ represents the 1-D weight kernel of the jth feature map of the ℓth layer while b_j^{ℓ} denotes the bias of the jth feature map of the ℓth layer, a non-linear activation function $f(\cdot)$, which in the most case would be a rectified linear unit activation (ReLu) [31].

3.2.2 Dilated Convolutions

For a 1-D sequence input x$\in \mathbb{R}^n$ and a filter $f : \{0, \ldots, k-1\} \rightarrow \mathbb{R}$, the dilated convolution operation F on element s of the sequence is defined as:

$$F(s) = (x *_d f)(s) = \sum_{i=1}^{k-1} f(i) \cdot x_{s-d \cdot i} \tag{6}$$

where d denotes the dilation factor, k is the filter size, and $s - d \cdot i$ denotes the direction of the past. As Fig. 2 shows, with the number of layers increases, dilation becomes larger, the output of the top layer can represent a larger range of inputs, which effectively expanding the receptive field of a ConvNet. The receptive field is crucial for time-series modeling because it explicitly limits the learnable feature periodicity at a given layer.

3.2.3 Residual Connections

A residual connection combines the input and the convolution signal of the layer (As shown in Fig. 3). This kind of learning framework enables the training process easier and it effectively allows the layer to learn the modification of the identity map which is beneficial to deep network training. Let $\hat{Z}_t^{(j,l)}$ be the result of the dilated convolution of the l_{th} layer and j_{th} block at time t and $Z_t{}^{(j,l)}$ be the result after adding the residual connection, denoted by

$$\hat{Z}_t^{(j,l)} = f(W_0 \hat{Z}_{t-d}^{(j,l-1)} + W_1 \hat{Z}_t^{(j,l-1)} + b) \tag{7}$$

and

$$Z_t{}^{(j,l)} = Z_t^{(j,l-1)} + V \hat{Z}_t^{(j,l)} + e \tag{8}$$

where $W_i \in \mathbb{R}^{F_w \times F_w}$ and weight matrices W = $[W_0, W_1]$ parameterizes the filter, F_w denotes the number of filters, b is the bias vector and b $\in \mathbb{R}^{F_w}$. $V \in \mathbb{R}^{F_w \times F_w}$ denotes the weight matrix and e $\in \mathbb{R}^{F_w}$ is the bias vector for the residual block.

As Fig. 3 shows, in a residual block, the TCN has two layers of dilated causal convolution and non-linearity. Weight normalization was applied to convolutional filters for normalization. Besides, for regularization, a spatial dropout was added after each dilated convolution.

However, in standard ResNet [32] the input is added directly to the output of the residual function, while in TCN (and general ConvNets), the input and output can have different widths. To account for the difference between the input and output widths, we use an additional 1×1 convolution to ensure element-wise addition $\bigoplus$ receiving tensors of the same shape.

The proposed model combined empirical mode decomposition with temporal convolutional network to predict RUL as shown in Fig. 4. First, The empirical mode decomposition is used to preprocess the data and obtain intrinsic mode functions

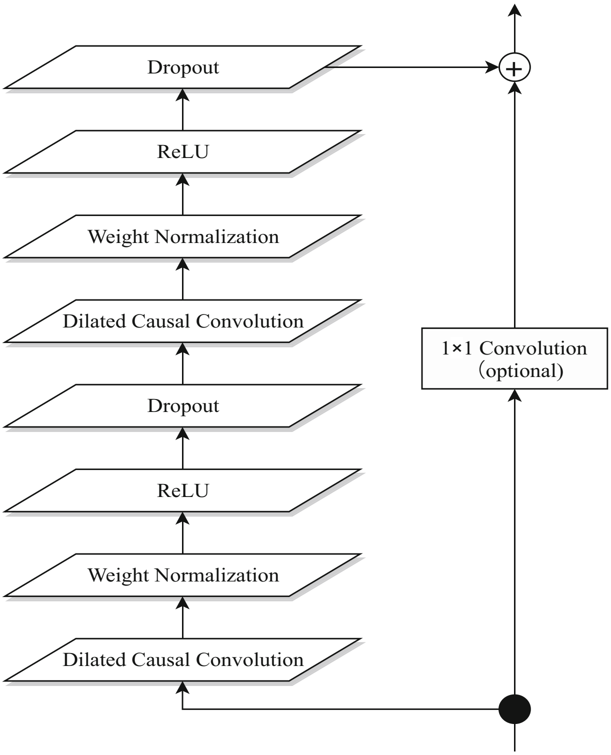


Fig. 3 A residual block of TCN

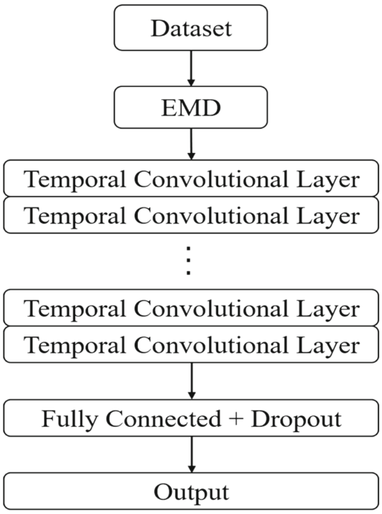


Fig. 4 The EMD–TCN framework for RUL estimtion

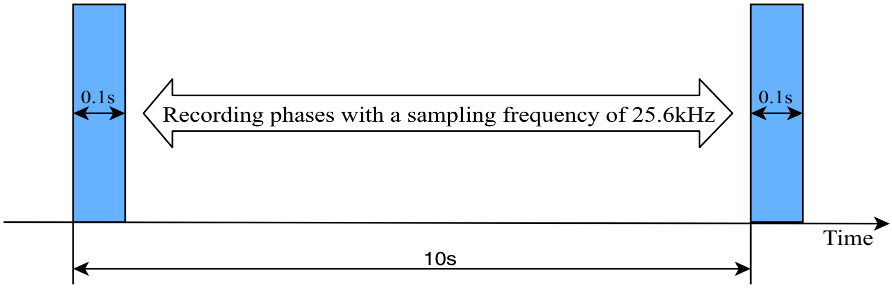


Fig. 5 Vibration signal data collection scheme

(IMFs) which represent various features of original signals in different time scales. Then, features of original signals are analyzed and the remaining useful life is predicted by TCN. EMD–TCN model has the capability of feature extraction like CNN and it is capable of building long term time dependencies like LSTM. The proposed framework can provide valuable information for enterprises to estimate the remaining useful life.

4 Experiments

This section is organized into three parts. Firstly, we introduce the dataset used in our RUL estimation. Secondly, we show the performance measures of our model. Finally, the results of our experiments were demonstrated.

4.1 Dataset Description

The PHM2012 dataset is widely used as a criterion of performance evaluation. It was collected from a new experimental platform for equipment bearing accelerated aging test, PRONOSTIA [11]. The platform provides aging data for equipment bearings under different working conditions with different motor speeds and radial forces, where the operational data is consistent with the normal degradation process of the bearing. In other words, the bearing is running from a completely new state to the fault occurs.

PRONOSTIA consists of three main parts: the rotating part, the fault generating part and the data acquiring part. Bearing aging is accomplished by a radial force generator acting on the ball bearing, and the aging data of the bearing is collected by a shock sensor and a temperature sensor. The vibration sensor consists of two mutually perpendicular accelerators with a sampling frequency of 25.6 kHz in every 10 s and 0.1 s sampling data once. Figure 5 shows the data collection scheme.

Bearing aging data in PRONOSTIA is composed of three working conditions: (1) 1800 rpm with 4000 N load; (2) 1650 rpm with 4200 N load; (3) 1500 rpm with the 5000 N load. There are 7534 samples in the training dataset and 13,965 samples in the testing dataset, and each time point contains 2560 vibration data.

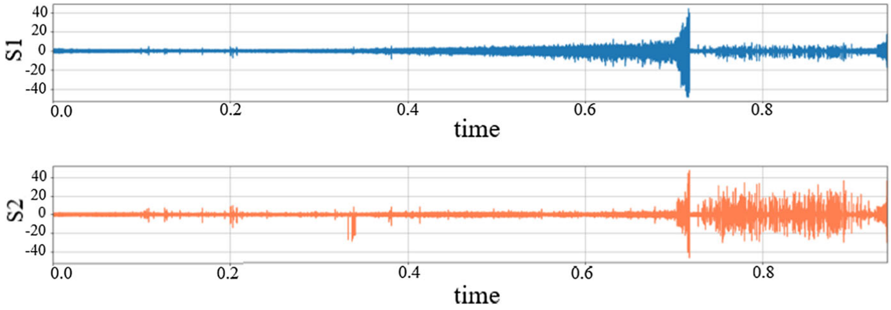


Fig. 6 Aging bearing data, two accelerator data for the first and second behavioral vibration sensors

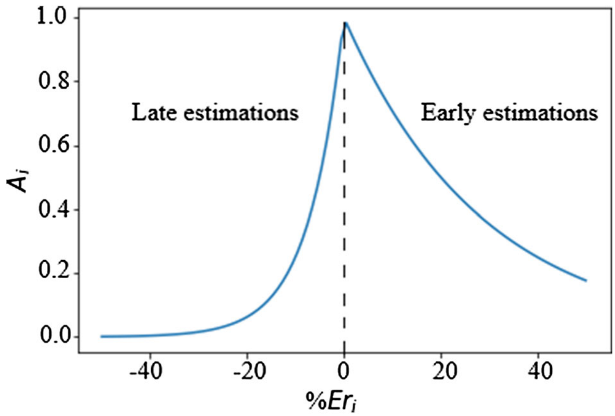


Fig. 7 The scoring function of RUL estimation

Depending on the bearing and its degradation process, the failure mode of each bearing varies significantly. As shown in Fig. 6, the full-life vibration data generated by PRONOSTIA under condition 1 for training is heterogeneous. The data generated by PRONOSTIA has the following aging modes: (a) The ideal aging mode. That is, as time progresses, the aging of the bearing with obvious and monotonous trends. Such data can easily predict the RUL of the device by using the threshold. (b) Sudden aging. In some cases, the aging of the bearing occurs abruptly without a slow incremental increase. (c) The theoretical model is mismatched. (d) The degree of noise affects the bearing aging process.

4.2 Performance Measures

The scoring function and mean square error (MSE) are used to evaluate the accuracy of the RUL estimation model. The remaining useful life of the bearing predicted by the model is RUL_i while $ActRUL_i$ represents the true remaining useful life of the bearing. The error rate of the ith test data is calculated by Eq. (9):

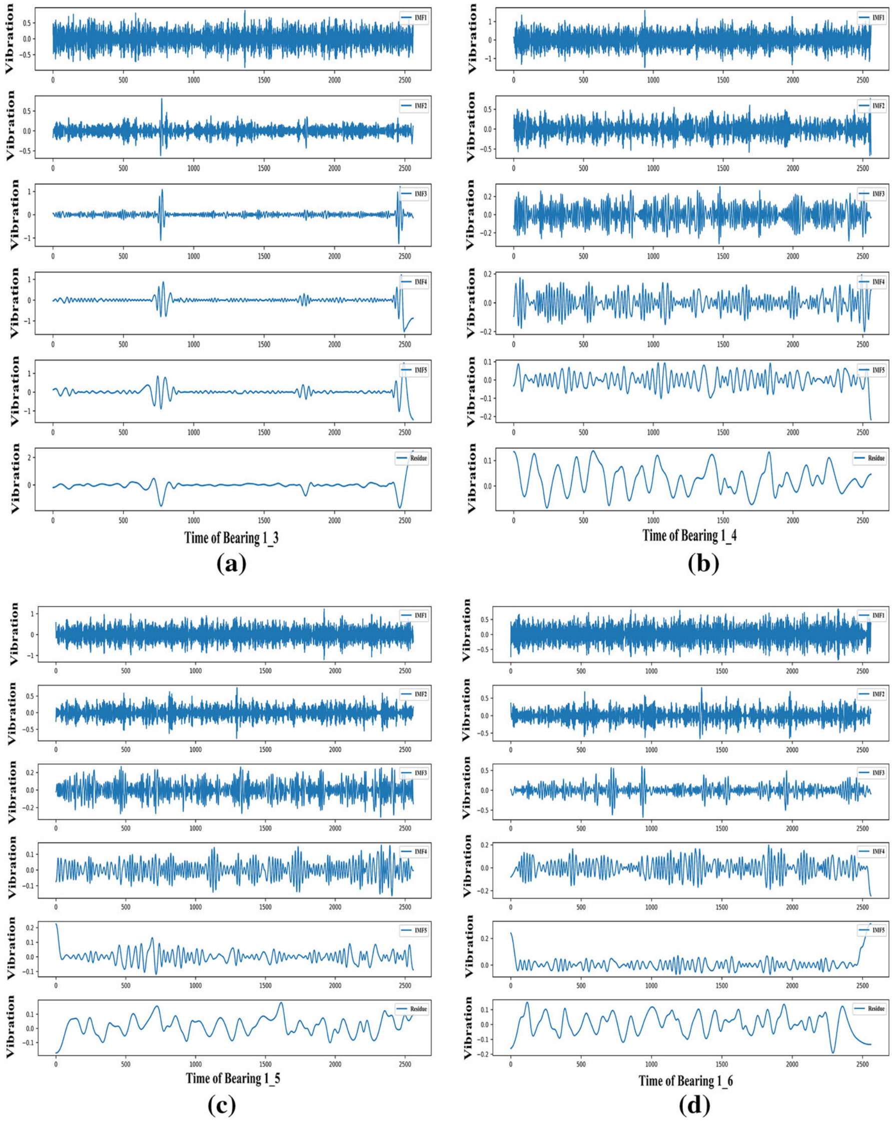
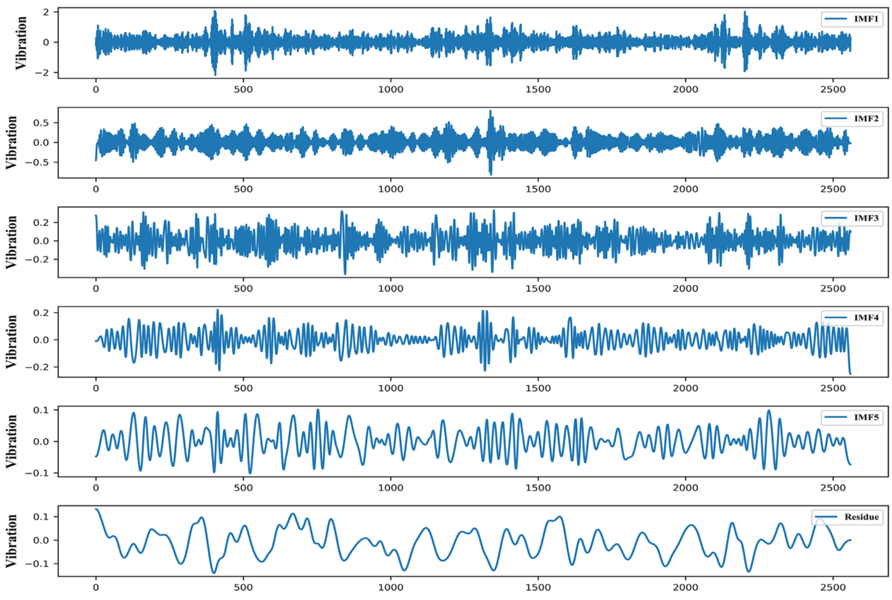


Fig. 8 EMD decomposition of accelerator sensor data



(e)

Fig. 8 continued

$$\%Er_i = 100 \times \frac{ActRUL_i - RUL_i}{ActRUL_i} \tag{9}$$

A positive rediction performance is the ability of the model to predict RUL earlier ($Er_i > 0$ or $RUL_i < ActRUL_i$), a negative prediction performance is that the model produces a higher prediction than the actual RUL ($Er_i < 0$ or $RUL_i > ActRUL_i$) which means failure will be occured to machine before the estimated time. Therefore, under-predictive and over-predictive will be treated in different forms, the accuracy score of RUL is calculated by Eqs. (10) and (11) :

$$A_i = \begin{cases} exp^{-ln(0.5)\cdot(Er_i/5)}, & \text{if } Er_i < 0 \\ exp^{+ln(0.5)\cdot(Er_i/20)}, & \text{if } Er_i \geq 0 \end{cases} \tag{10}$$

$$Score = \frac{1}{n}\sum_{i=1}^{n} A_i \tag{11}$$

Figure7 depicts the evolution of the scoring function. We also use Mean Square Error (MSE) to overcome the sensitivity to outliers of the scoring function. MSE is given by

$$MSE = \frac{\sum_{i=1}^{n} Er_i^2}{n} \tag{12}$$

Table 1 Test results of different convolutional neural networks

	*Bearing*1_3		*Bearing*1_4		*Bearing*1_5		*Bearing*1_6		*Bearing*1_7		*Total*	
	Score	*MSE*	*Score*	*MSE*	*Score*	*MSE*	*Score*	*MSE*	*Score*	*MSE*	*Score*	*MSE*
CNN	0.40	0.12	0.23	0.24	0.34	0.16	0.35	0.13	0.32	0.17	0.34	0.15
LSTM	0.35	0.16	0.30	0.15	0.30	0.20	0.22	0.22	0.21	0.23	0.27	0.19
TCN	0.61	**0.03**	0.36	0.15	0.44	0.10	0.44	0.10	0.46	0.12	0.44	0.11
EMD–TCN	**0.60**	0.06	**0.57**	**0.08**	**0.50**	**0.10**	**0.44**	**0.10**	**0.46**	**0.11**	**0.49**	**0.09**

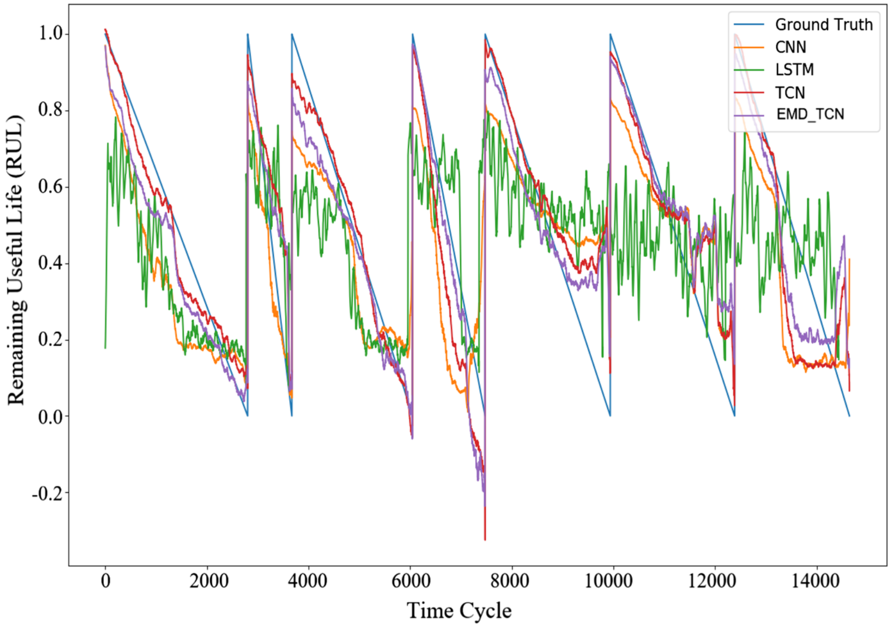


Fig. 9 The prediction results of different convolutional neural networks

4.3 Results and Discussion

The overall operation of the experiment is divided into three parts, decomposing the original input data, training with the TCN networks, and integrating the results to produce the final result.

The PHM 2012 data contains X-axis vibration and Y-axis vibration. First, EMD is used to decompose X and Y respectively, where five IMFs and the remaining residual are respectively obtained. Since there are 2560 vibration data at each time point, the original data mode is (2560, 2). After processed by EMD, the data mode is (2560, 12). The data that processed by EMD is shown in Fig. 8. As Fig. 8 shows, each IMF has different characteristics of fluctuating. The first several IMFs have more energy and possess more intrinsic information than the latter.

We trained TCN model by using processed bearing aging training dataset, then, test our framework on testing dataset. We conduct a series of comprehensive experiments to manifest the superiority of EMD–TCN framework, CNN and LSTM are investigated for comparisons. The experimental results are shown in Table 1. The results indicate that LSTM and CNN can hardly capture degrading trend exactly but TCN is effective in prediction by combine history condition and convolution. Our framework intergrated TCN with EMD method has a significant improvement compared with the traditional convolutional neural network. In this paper, we have shown that the average score of EMD–TCN model is improved by 10–20% than CNN and LSTM and the error of EMD–TCN model is minimum, which proves that our method is effective to remaning useful life estimation.

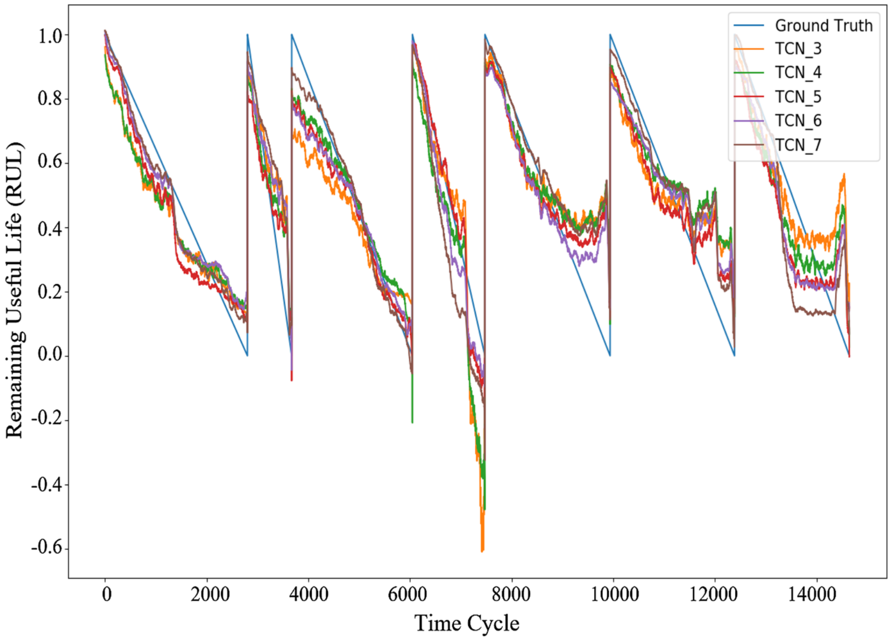


Fig. 10 The prediction results of TCN with different dilation value

Table 2 Sensitive test of different dilation value of TCN

	TCN_3	TCN_4	TCN_5	TCN_6	TCN_7
Score	0.41	0.42	0.46	0.47	0.47
MSE	0.12	0.11	0.10	0.10	0.10

The prediction results of different convolutional neural networks are compared with the ground truth of the RUL have shown in Fig. 9. All of networks can detect the degradation situation on the test bearings and all predict curves are downtrend during the lifecycle of bearings. Figure 9 depicts that the result of EMD–TCN framework is closest to the ground truth, EMD–TCN framework outperformed in all convolutional neural networks.

The influence of the number of dilation of TCN on the prognostic performance is investigated and the results are presented in Fig. 10. The value of score and MSE are described in Table 2, Our results indicate that with the increase of the dilation value, the prediction accuracy has increased initially and then remains the same. The training loss of TCN with different dilation value is depicted in Fig. 11. With the increase of dilation value, the loss of TCN model is decreased and the TCN model has a better generalization ability. The TCN with a higher dilation value is more robust for remaining useful life estimation.

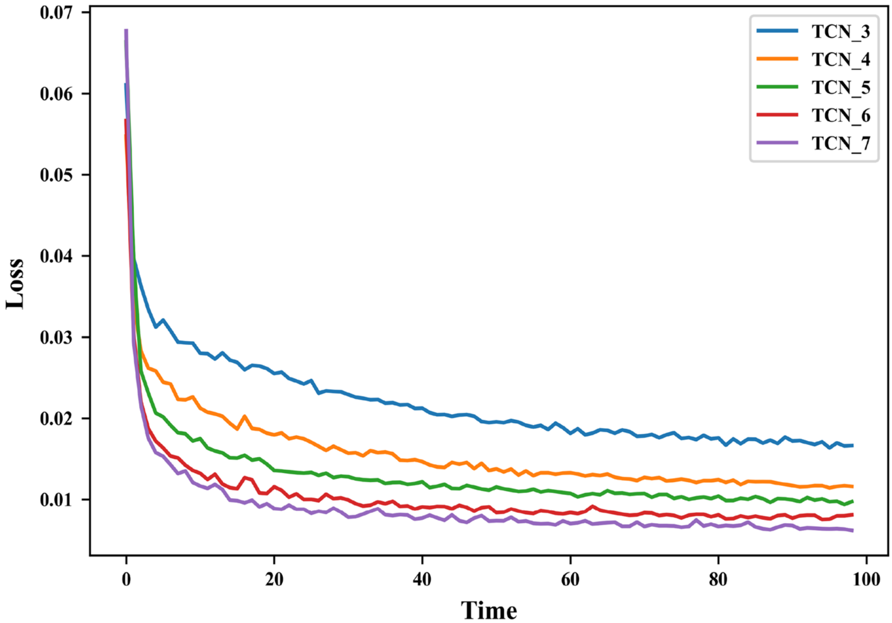


Fig. 11 The loss of TCN with different dilation value

5 Conclusion

This paper proposed a novel deep learning model structure for RUL estimation. We firstly decompose the original signal data by empirical mode decomposition (EMD) and expand the data to 12 dimensions. Then, the processed datasets are used to train the TCN respectively. The PHM2012 dataset is used to contrast the performance of different models. Experiments are conducted to demonstrate the effectiveness and superiority of our EMD–TCN model. The experimental results have shown that the EMD–TCN method is effective for RUL estimation of industrial applications. Finally, the impact of memory length has been explored. In the future, we plan to extend the proposed framework to other prognostic applications.

Acknowledgements This work is supported by the National Key R&D Program of China (No. 2018YFB1004804), National Natural Science Foundation of China (No. 61702492), Shenzhen Discipline Construction Project for Urban Computing and Data Intelligence, and Shenzhen Basic Research Program (No. JCYJ20170818153016513).

References

1. Heng, A., Tan, A.C., Mathew, J., Montgomery, N., Banjevic, D., Jardine, A.K.: Intelligent condition-based prediction of machinery reliability. Mech. Syst. Signal Process. **23**(5), 1600–1614 (2009)
2. Zhang, J., Jiang, N., Li, H., Li, N.: Online health assessment of wind turbine based on operational condition recognition. Trans. Inst. Meas. Control. **41**(10), 2970–2981 (2019)

3. Fan, M., Zeng, Z., Zio, E., Kang, R., Chen, Y.: A sequential bayesian approach for remaining useful life prediction of dependent competing failure processes. IEEE Trans. Reliab. **68**(1), 317–329 (2018)
4. Khorasgani, H., Kulkarni, C., Biswas, G., Celaya, J.R., Goebel, K.: Degredation modeling and remaining useful life prediction of electrolytic capacitors under thermal overstress condition using particle filters. In: Annual Conference of the Prognostics and Health Management Society (PHM'13) (2013)
5. Zheng, Z., Podobnik, B., Feng, L., Li, B.: Changes in cross-correlations as an indicator for systemic risk. Sci Rep **2**, 888 (2012)
6. Zhang, C., Hong, G.S., Xu, H., Tan, K.C., Zhou, J.H., Chan, H.L., Li, H.: A data-driven prognostics framework for tool remaining useful life estimation in tool condition monitoring. In 2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), pp. 1–8. IEEE (2017)
7. Si, X.-S., Wang, W., Hu, C.-H., Zhou, D.-H.: Remaining useful life estimation—a review on the statistical data driven approaches. Eur. J. Oper. Res. **213**(1), 1–14 (2011)
8. Fu, D., Zheng, Z., Gui, J., Xiao, R., Huang, G., Li, Y.: Development of a fuel management model for a multi-source district heating system under multi-uncertainty and multi-dimensional constraints. Energy Convers. Manag. **153**, 243–256 (2017)
9. Li, X., Ding, Q., Sun, J.-Q.: Remaining useful life estimation in prognostics using deep convolution neural networks. Reliab. Eng. Syst. Saf. **172**, 1–11 (2018)
10. Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., Liu, H.H.: The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **454**(1971), 903–995 (1998)
11. Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-Morello, B., Zerhouni, N., Varnier, C.: Pronostia: an experimental platform for bearings accelerated degradation tests. In: IEEE International Conference on Prognostics and Health Management, PHM'12. IEEE Catalog Number: CPF12PHM-CDR, pp. 1–8 (2012)
12. Ahmadzadeh, F., Lundberg, J.: Remaining useful life estimation. Int. J. Syst. Assur. Eng. Manag. **5**(4), 461–474 (2014)
13. Gao, Z., Cecati, C., Ding, S.X.: A survey of fault diagnosis and fault-tolerant techniques—part I: fault diagnosis with model-based and signal-based approaches. IEEE Trans. Ind. Electron. **62**(6), 3757–3767 (2015)
14. Luo, J., Bixby, A., Pattipati, K., Qiao, L., Kawamoto, M., Chigusa, S.: An interacting multiple model approach to model-based prognostics. In: SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance (Cat. No. 03CH37483), vol. 1, pp. 189–194. IEEE (2003)
15. Li, Y., Kurfess, T., Liang, S.: Stochastic prognostics for rolling element bearings. Mech. Syst. Signal Process. **14**(5), 747–762 (2000)
16. Baillie, D., Mathew, J.: Nonlinear model-based fault diagnosis of bearings. NDT e Int. **5**(30), 328 (1997)
17. Tobon-Mejia, D.A., Medjaher, K., Zerhouni, N., Tripot, G.: A data-driven failure prognostics method based on mixture of gaussians hidden markov models. IEEE Trans. Reliab. **61**(2), 491–503 (2012)
18. Cheng, F., Qu, L., Qiao, W., Hao, L.: Enhanced particle filtering for bearing remaining useful life prediction of wind turbine drivetrain gearboxes. IEEE Trans. Ind. Electron. **66**(6), 4738–4748 (2018)
19. Banjevic, D., Jardine, A.: Calculation of reliability function and remaining useful life for a markov failure time process. IMA J. Manag. Math. **17**(2), 115–130 (2006)
20. Sankararaman, S., Goebel, K.: Uncertainty quantification in remaining useful life of aerospace components using state space models and inverse form. In: 54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, p. 1537 (2013)
21. Zhang, J., Wang, S., Chen, L., Guo, G., Chen, R., Vanasse, A.: Time-dependent survival neural network for remaining useful life prediction. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 441–452. Springer (2019)
22. Tian, Z.: An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring. J. Intell. Manuf. **23**(2), 227–237 (2012)
23. Soualhi, A., Medjaher, K., Zerhouni, N.: Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression. IEEE Trans. Instrum. Meas. **64**(1), 52–62 (2014)
24. Zhu, M., Ye, K., Wang, Y., Xu, C.-Z.: A deep learning approach for network anomaly detection based on AMF-LSTM. In: IFIP International Conference on Network and Parallel Computing, pp. 137–141. Springer (2018)

25. Zhu, M., Ye, K., Xu, C.-Z.: Network anomaly detection and identification based on deep learning methods. In: International Conference on Cloud Computing, pp. 219–234. Springer (2018)
26. Lin, P., Ye, K., Xu, C.-Z.: Dynamic network anomaly detection system by using deep learning techniques. In: International Conference on Cloud Computing, pp. 161–176. Springer (2019)
27. Ye, K., Liu, Y., Xu, G., Xu, C.-Z.: Fault injection and detection for artificial intelligence applications in container-based clouds. In: International Conference on Cloud Computing, pp. 112–127. Springer (2018)
28. Pinheiro, H., Jain, P., Joos, G.: Series-parallel resonant converter in the self-sustained oscillating mode for unity power factor applications. In: Proceedings of APEC 97-Applied Power Electronics Conference, vol. 1, pp. 477–483. IEEE (1997)
29. Bai, S., Kolter, J. Z., Koltun, V.: An empirical evaluation of generic convolutional and recurrent networks for sequence modeling (2018). arXiv preprint arXiv:1803.01271
30. Deng, S., Zhang, N., Zhang, W., Chen, J., Pan, J. Z., Chen, H.: Knowledge-driven stock trend prediction and explanation via temporal convolutional network. In: Companion Proceedings of The 2019 World Wide Web Conference, pp. 678–685. ACM (2019)
31. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th international conference on machine learning (ICML-10), pp. 807–814 (2010)
32. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778 (2016)