Reliability Engineering and System Safety 225 (2022) 108590

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Trend attention fully convolutional network for remaining useful life estimation

Linchuan Fan [a,b], Yi Chai [a,b,*], Xiaolong Chen [b]

[a] *Key Laboratory of Complex System Safety and Control, Ministry of Education, No. 174, Shazheng Street, Shapingba District, Chongqing 400044, China*
[b] *College of Automation, Chongqing University, No. 174, Shazheng Street, Shapingba District, Chongqing 400044, China*

ARTICLE INFO

Keywords:
Remaining useful life
Data-driven prognostic
Signal selection
Attention mechanism
Interpretability

ABSTRACT

Modern engineered systems usually employ multiple sensors to monitor equipment health status. However, most remaining useful life (RUL) estimation methods based on deep learning are hard to select helpful signals and remove useless signals accurately. Moreover, the attention mechanisms they employed could hardly obtain an optimal attention distribution at an acceptable computational cost, resulting in poor prediction performance. Therefore, we proposed a novel signal selection method, terming the "Loss boundary to Mapping ability" (LM) approach. It can accurately select the signals that can contribute to RUL prediction tasks. Then, inspired by the characteristics of RUL monitoring signals, we proposed a novel end-to-end framework called Trend attention Fully Convolutional Network (TaFCN) to enhance prediction performance further. These two methods constitute our prognostic method. We conducted a series of ablation experiments and comparative experiments with recent methods on the C-MAPSS turbofan engine dataset. The ablation experiments proved the necessity and advanced performance of the LM and the proposed attention mechanism employed in the TaFCN. The comparative experiments demonstrated the state-of-the-art performance of our prognostic method. Furthermore, we developed an interpretability analysis method, which revealed the logical reasoning process of our method.

1. Introduction

To guarantee the reliable running of industrial activities, modern industry managers usually need to employ the Prognostics and Health Management system (PHM) to determine the health status of equipment or crucial components. As a vital task of the PHM system, Condition-Based Maintenance (CBM) includes diagnosis and prediction. Diagnosis is to locate the reason for equipment failure. Prediction is to estimate the remaining time until the failure occurrence, that is, predicting Remaining Useful Life (RUL). To avoid fateful consequences resulting from the sudden failure of running industrial equipment, Maintenance workers must get the accurate RUL of equipment or crucial components to make maintenance decisions [1]. Therefore, RUL prediction is a critical task in the PHM system.

Generally, existing RUL prediction methods can be roughly grouped into three categories, i.e., model-based method, data-driven method, and hybrid method [2]. The model-based approach needs to establish the corresponding degradation mechanism model according to the extensive prior knowledge of the specific object. However, with the increase of equipment structure complexity, it is troublesome to establish degradation mechanism models of system objects, which makes model-based methods highly limited. The data-driven approach employs machine learning or probability statistics to reveal the underlying correlations and causalities between signal data and RUL label [1]. Since it does not require a degradation mechanism model, it is easy to apply to actual industrial activities and has become the most popular RUL prediction method. The hybrid method combines the degradation mechanism model with the data-driven approach to predict equipment RUL [3]. Unfortunately, the combination approach and the high cost of obtaining an accurate degradation mechanism model put the hybrid method into trouble. Hence, this paper will focus on the data-driven approach.

The data-driven method aims to adaptively establish an approximate model of degradation mechanism model based on signal data and RUL label. As shallow machine learning models, Auto-regressive Model [4], Wiener-process [5], and Bayesian approach [6] have been widely used in RUL prediction tasks. Most of these methods require manual extraction of data features. Its prediction performance largely depends on feature quality. And its heavy feature engineering dramatically increases the complexity of RUL prediction tasks. Fortunately, the deep learning model significantly reduces the complexity

[*] Corresponding author at: College of Automation, Chongqing University, No. 174, Shazheng Street, Shapingba District, Chongqing 400044, China.
E-mail address: chaiyi@cqu.edu.cn (Y. Chai).

https://doi.org/10.1016/j.ress.2022.108590
Received 29 December 2021; Received in revised form 9 May 2022; Accepted 10 May 2022
Available online 18 May 2022

of tasks through end-to-end mapping and achieves promising results through powerful arbitrary function fitting capability. These characteristics have attracted the attention of numerous researchers.

Recently, Convolutional Neural Networks (CNN), born in image fields, have been introduced into RUL prediction tasks. Contributed by weight sharing and translation invariance, CNN has excellent high-level abstract representation capability. Li et al. proposed a deep convolution neural network (DCNN) for RUL prediction, which employed five convolutional layers to extract high-level abstract representations of sensor data [4]. Li et al. used convolutional layers containing filters of different sizes to extract the multi-scale features of the time-spectrogram converted from the vibration data sequence and achieved better than a single scale [7]. Yang et al. proposed a dual CNN architecture, which first uses CNN to identify the early points of faults, then uses CNN and a fully connected layer to fit the mapping model between reliability variable and RUL value [8]. Cao et al. proposed a temporal convolutional network and residual self-attention mechanism. After original signals are transformed into a marginal spectrum, the method can capture long short-term dependencies and extract deep high-level features [9]. Peng et al. proposed a spatial–temporal feature fusion method that uses CNN and LSTM to extract the spatial–temporal features [10]. Zhuang et al. proposed a cross-domain approach that employed a temporal convolution network to extract fruitful degradation information. Then a cross-domain adaption architecture was utilized to capture domain invariant information [11].

As a variant of CNN, Fully Convolutional Network (FCN) has attracted considerable attention from time series analysis researchers. A large number of experiments have verified that the performance of FCN far exceeds LSTM in the task of time series data classification [12–14]. This shows that FCN has powerful feature extraction capabilities for time series data. Therefore, we will employ FCN to extract high-level abstract feature representations for RUL estimation.

Signal selection is essential for the performance of RUL prediction tasks. Aiming at selecting the appropriate signal, some work has been done. In [10], researchers selected all signals as input to the model, which makes the input contain all available RUL information. However, the interference of irrelevant information may increase the computational cost and model complexity, thereby reducing the model performance. To select signals rich in RUL information, numerous researchers select signals whose value changes monotonously over the running time [4,15–18]. Since device RUL is a linear function of the running time, these signals have excellent mapping capability to the RUL. However, some signals with irregular trends may contain critical RUL information in a specific life cycle. This approach fails to select these signals, thus unsatisfactory prediction performance. Wu et al. calculated a weighted indicator by the monotonicity and correlation between signal value and time, then select the signal with a high indicator value by threshold [19]. However, the sequence obtained by an arbitrary linear function can get highly large monotonicity and correlation values. Hence, some signals that are not related to RUL may be selected. To solve the above problems, we proposed a novel loss-boundary-value-based signal selection method, which can accurately retain useful signals and eliminate useless signals for prediction tasks.

For the RUL prediction task, the importance of signals is probable to be different. Therefore, attention mechanism is fundamental to the model performance. Attention mechanism originated in the image domain. SeNet is one of the most crucial attention mechanisms. Its Squeeze-and-Excitation (SE) block can effectively explore the correlation between image channels and significantly improve model performance [20]. Researchers have carried out much research on SeNet. For example, Artacho et al. employed SeNet to compute the importance weights of backbones in human poses, proposing UniPose+, a framework for robust and efficient 2D and 3D human pose estimation [21]. To solve speech emotion modeling in a long-term context, Zhao et al. proposed a discrete speech emotion recognition neural network framework integrating parallel convolutional layers and SeNet to calculate the channel importance of the paralleled convolution feature [22].

Unlike the fruitful attention mechanism research in the image field, the attention mechanism research for RUL prediction tasks has just started. In [16], the researcher directly sends all data points to a fully connected layer to calculate the importance of each signal. This method uses all data points of signal samples to calculate signal importance, where sufficient information is helpful to the mining of correlation between signals to a certain extent. However, it dramatically increases the computational cost and may cause parameters redundancy of fully connected layer and reduce model performance. In [15,23,24], researchers first calculate the importance of each signal sample separately, then uses the softmax function to standardize signal importance. This way can reduce the computational cost and parameter redundancy to a certain extent. However, it inhibits the correlation mining between different signals. Hence, this method is probable to cause the output value to deviate from the ideal importance significantly. All the above methods use all the points of a signal sample to calculate signal importance. They do not design efficient representation method for the signal sample.

Under the methods mentioned above, it is impossible to fully explore the signal correlation and obtain the optimal attention distribution in non-parametric redundancy. To solve the above problems, we propose an attention mechanism, namely trend attention network (TaNet), which utilizes the trend characteristics of RUL monitoring signals to characterize each signal sample effectively, saving computational costs and fully exploring the correlation between signals.

This paper contributes a novel prognostic method and an interpretability analysis method for RUL prediction tasks. First, we proposed the Loss boundary to Mapping ability (LM) signal selection method, which can quantify the ability of signals to map equipment RUL, thus selecting helpful signals and removing useless signals accurately. Then, inspired by signal characteristics, we proposed a novel end-to-end framework called Trend attention Fully Convolutional Network (TaFCN) for RUL prediction tasks, which extracts trend information to represent samples efficiently, calculates attention distribution by excitation operation, achieving optimal attention distribution at a small computational cost, extracts deep features by fully convolutional layers and calculates equipment RUL. Finally, we developed modified Class Activation Mapping (CAM) for the interpretability analysis of RUL prediction tasks. We employed the weights of the last two fully connected layers to calculate the weight coefficients of each channel of the last convolutional layer, then calculated the importance of the activation at all temporal locations.

The main contributions are listed as follows:

(1) The proposed signal selection method LM can select valuable signals and remove useless signals accurately by quantifying the ability of signals to map equipment RUL.
(2) We proposed an efficient and powerful attention mechanism called trend attention network (TaNet), which is specially designed by the monitoring signals characteristics of RUL prediction tasks.
(3) Based on TaNet, we proposed a novel end-to-end framework called Trend attention Fully Convolutional Network (TaFCN) for RUL prediction tasks.
(4) We developed an interpretability analysis method for RUL prediction tasks based on the interpretability analysis method designed for images. To our best knowledge, this is the first time that the interpretability analysis method was developed for RUL prediction tasks.
(5) A series of ablation experiments, comparative experiments, and interpretability analysis experiments were conducted to demonstrate the necessity and advanced performance of LM and TaNet, prove the superiority of the prognostic method, and figure out the logical reasoning of the prognostic method, respectively.

The remainder of this paper is outlined as follows. We start with our prognostic method in Section 2. Next, a series of ablation experiments and comparison experiments are presented in Section 3. Then, we analyzed the interpretability of our proposed methods in Section 4. Finally, Section 5 concludes this paper and gives perspectives for future work.

2. Proposed method

The proposed prognostic method consists of two parts: Loss boundary to Mapping ability (LM) signal selection method and Trend attention Fully Convolutional Network (TaFCN).

2.1. Formulation of multivariate RUL prediction

For multivariate RUL prediction tasks, the train set is denoted as $S^{tr} = \{(x_1^{tr}, y_1^{tr}), \dots (x_r^{tr}, y_r^{tr}), \dots (x_{N_{tr}}^{tr}, y_{N_{tr}}^{tr})\}$, where N_{tr}, $x_r^{tr} \in \mathbb{R}^{C\times W}$, and y_r^{tr} refer to the sample number of train set, the rth input samples in train set, and the RUL labels of the rth input samples in train set, respectively. C is the number of RUL monitoring signals, W is the size of the time window. The test set is denoted as $S^{te} = \{(x_1^{te}, y_1^{te}), \dots (x_e^{te}, y_e^{te}), \dots (x_{N_{te}}^{te}, y_{N_{te}}^{te})\}$, where N_{te}, $x_e^{tr} \in \mathbb{R}^{C\times W}$, and y_e^{tr} are the sample number of test set, the eth the input sample in test set, and the RUL label of the eth input sample in test set, respectively.

The multivariate RUL prediction task is to learn a remaining life prediction model $f(x)$ by the given S^{tr}, so that all samples in S^{te} satisfy $y_e^{te} = f(x_e^{te})$ as much as possible.

2.2. Loss boundary to mapping ability

As the first step of multivariate RUL prediction methods, Signal selection is a vital data preprocessing process on the performance of RUL prediction. To improve the performance of RUL prediction tasks and reduce the complexity of learning tasks, we need to retain valuable signals and remove useless signals accurately. Hence, we propose the LM approach, which is a general signal selection method not limited to monitoring signals for RUL prediction tasks.

In LM, we proposed the loss boundary value $loss_b$ for judging whether a signal is helpful to the prediction task. It is given by:

$$loss_b = \min\left(\sqrt{\frac{1}{N_s}\sum_{j=1}^{N_s}(x - y_{true}^j)^2}\right) \tag{1}$$

where N_s, x, and y_{true}^j is the training samples number of the signal, an independent variable, and the label of the jth training sample, respectively.

When a signal does not contain any RUL information, each training sample obtained by the signal cannot effectively map the corresponding RUL through its sample information. This means that any model cannot reduce the training loss value by establishing the actual mapping relationship between RUL and input samples. At this time, the output value of the model has two cases. One is that each output value is a random value, and the other is that all output values converge to a value, which is the local or global optimal point of the loss function obtained by the true label distribution of training samples. Corresponding to the global optimal point of the training model, the $loss_b$ is the minimum loss value expected by the training model when the training samples do not contain any information related to RUL. Hence, we employed $loss_b$ as the threshold value for judging whether the signal data contains information contributing to prediction tasks.

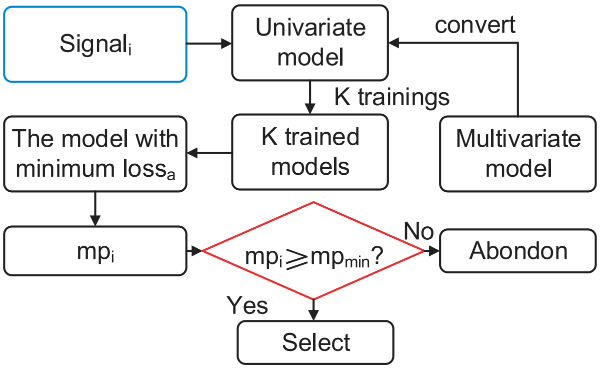


Fig. 1. Flowchart of LM signal selection method.

The flowchart of LM is shown in Fig. 1. $Signal_i$ represents the data obtained by the ith signal. K, $loss_a$, and mp_{min} is the number of model training, the average loss value of the last 10 epochs in training, and the minimum value of mp, respectively. mp_i represents the ability of the ith signal to map RUL, is given by:

$$mp_i = \mathrm{Relu}\left(\frac{loss_b - loss_a}{loss_b}\right) \tag{2}$$

where Relu refers to rectified linear unit [25], the range of mp_i is $[0, 1)$, the larger the mp_i, the stronger the ability of the ith signal data to map RUL. mp_i equal to 0 means that the ith signal data is useless for RUL prediction tasks.

In the flowchart of LM, first, we use Eq. (1) to calculate the loss boundary value $loss_b$. Even if training samples do not contain any information related to RUL, the model can rely on the true label distribution of the training samples to converge all predicted values to x in Eq. (1) to reduce the loss value during training. Hence, when the training samples do not contain any RUL information and the training model relies on the true label distribution of the training sample to predict x in Eq. (1) to reduce loss value, $loss_b$ is the minimum loss value expected by the training model. Then, we convert the multivariate prediction model to a univariate prediction model and train the converted univariate prediction model with the training set data of the ith signal. After K trainings, we take the model with minimum $loss_a$, which is the average loss value of the last ten epochs in training. The setting of K trainings is to prevent an accidental phenomenon in the neural network that training set loss ceases change with a highly massive value in the first epoch of training. Because this situation is accidental and the probability is small, in our experiments, K is set to 3. Then, $loss_a$ and $loss_b$ are employed to calculate mp_i, which represents the ability of the ith signal to map RUL. Eq. (2) can be understood as when $loss_a$ is within the boundary ($loss_a < loss_b$), the farther $loss_a$ is from the boundary $loss_b$, the larger mp_i is. When mp_i is greater than the ability to map RUL minimum threshold mp_{min} we set, the ith signal is selected for our RUL prediction task. To adopt as many signals as possible that are associated with RUL prediction tasks, mp_{min} was set to 0 in our experiments.

2.3. Trend attention fully convolutional network

The attention mechanism in the neural network is the same as human visual attention. It can locate crucial targets, then put more attention to them. This means that the attention mechanism can effectively choose the informative segment from data, thereby remarkably improving task performance.

To quantify the importance of different signals for the RUL prediction tasks, we proposed an attention mechanism, namely, trend attention network (TaNet).

As shown in Fig. 2, TaNet converts the input $x \in \mathbb{R}^{C\times H\times 1}$ into the feature tensor $u \in \mathbb{R}^{C\times H\times 1}$, then $u \in \mathbb{R}^{C\times H\times 1}$ is employed as the input of FCN. Mathematically, the tensor $g \in \mathbb{R}^{C\times N\times 1}$ attained by trend squeeze operation can be expressed as

$$g(c, n) = \frac{1}{floor(H/N)}\sum_{i=n*floor(H/N)}^{(n+1)*floor(H/N)} x(c, i) \tag{3}$$

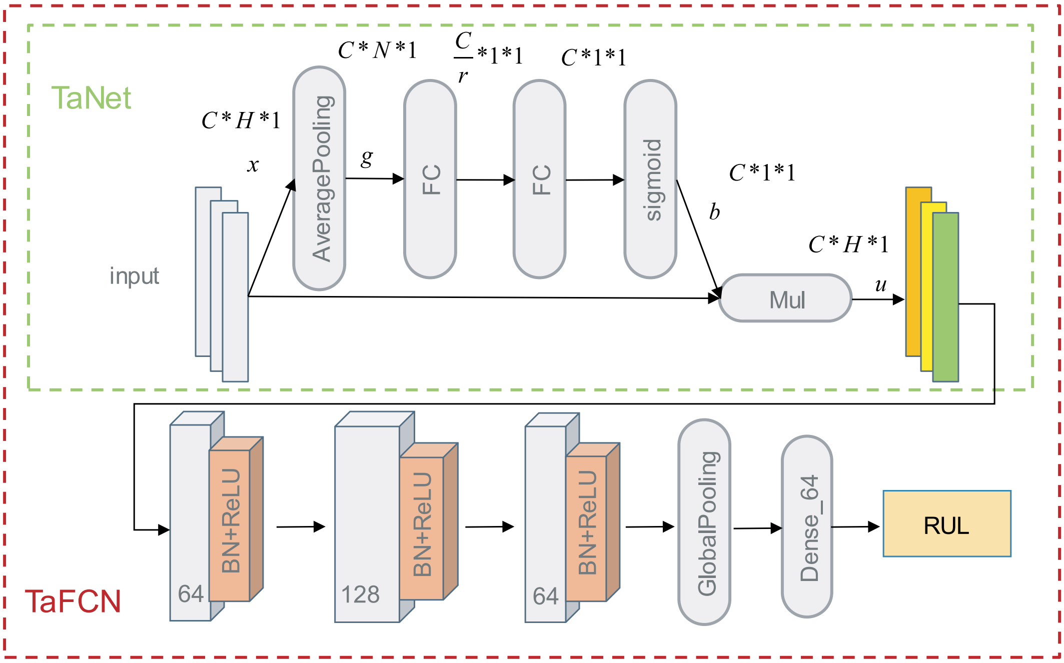


Fig. 2. Trend attention Fully Convolutional Network (TaFCN).

where $floor(x)$ is the function which returns the largest integer no bigger than the input x. H and $x(c,i)$ refer to the length of the input feature and the ith element of the cth channel of inputs. C is the number of the signals selected by LM. N is the length of the vector representing the trend, which not only affects the complexity of TaNet, but also affects its performance. It is worth noting that the trend squeeze operation of TaNet is different from the squeeze operation of SeNet. The piecewise mean value attained by the trend squeeze operation of TaNet is more informative than the global mean of the sequence attained by the squeeze operation of SeNet. More importantly, the global mean completely loses the rising and falling trends information of the sequence. The piecewise mean value attained by the trend squeeze operation can efficiently characterize the trend information of signal samples. Due to the role of the trend squeeze operation in characterizing the changing trend of the series, TaNet is also suitable for other multivariate time series data with changing trends.

The excitation vector $b \in \mathbb{R}^{C\times 1\times 1}$ is described as

$$b = \xi(W_2\sigma(W_1 g)) \tag{4}$$

where $\sigma, \xi, W_1 \in \mathbb{R}^{\frac{C}{r}\times C}$, and $W_2 \in \mathbb{R}^{C\times\frac{C}{r}}$ refer to rectified linear unit, sigmoid activation function, the parameters of the first fully connected layer (FC), and the parameters of the second FC, respectively. r is used for adjusting the complexity of TaNet [20].

The output of TaNet is calculated by

$$u = Mul(b, x) \tag{5}$$

where $Mul(g,x)$ is the channel-wise multiply function. After trend squeeze operation Eq. (3), the trend characteristics of RUL monitoring signals were extracted to represent each signal sample effectively. Then, the excitation operation containing two fully connected layers and a sigmoid activation function was employed to calculate importance weights for all signals. The input sample containing multiple signal samples is multiplied by the signal's importance weight by a channel-wise multiply function Eq. (5) to obtain a new weighted input sample. Finally, we feed the new weighted input samples into the next network structure in TaFCN. In this way, we achieve the importance representation in the network.

To balance the complexity and performance of the model, N and r were set to 3 and 1 in our experiments respectively. Due to the number of network neurons input to the fully connected layer in TaNet has been dramatically reduced after trend squeeze operation, it is not necessary to increase the r of Eq. (4) to reduce the number of network training of the TaNet, r is set to 1. The setting of N will be analyzed in Section 4. The effectiveness of TaNet will be proven in Section 3.2.2.

Then, fully convolutional network was employed to extract the deep features of signal data for RUL prediction [13,14,26].

Three fully convolutional layers are given by:

$$c_1 = \mathrm{Relu}(\mathrm{BN}(\mathbf{W}_{c_1} \otimes u + \mathbf{b}_{c_1})) \tag{6}$$

$$c_2 = \mathrm{Relu}(\mathrm{BN}(\mathbf{W}_{c_2} \otimes c_1 + \mathbf{b}_{c_2})) \tag{7}$$

$$c_3 = \mathrm{Relu}(\mathrm{BN}(\mathbf{W}_{c_3} \otimes c_2 + \mathbf{b}_{c_3})) \tag{8}$$

where $\otimes$, BN, and $Re\,lu$ refer to the convolution operator, batch normalization, and relu layer, respectively. W_{c_1}, b_{c_1}, W_{c_2}, b_{c_2}, W_{c_3}, and b_{c_3} refer to the parameters of the convolution kernel of the first fully convolutional layer, the biases of the first fully convolutional layer, the parameters of the convolution kernel of the second fully convolutional layer, the biases of the second fully convolutional layer, the parameters of the convolution kernel of the third fully convolutional layer, and the biases of the third fully convolutional layer, respectively. Batch normalization enables the value of each layer to be transferred to the next layer within the effective range and plays a role in preventing over-fitting. Hence, we did not add dropout and max pooling operation to prevent overfitting. For best prediction performance, the three filter sizes in convolutional layers were set to 64, 128, and 64, respectively. The three 1-D kernel sizes were set to 16, 10, and 6, respectively. The units of the dense layer after globe average pooling were set to 64.

Following the flowchart in Fig. 2, we receive the deep features extracted from the last fully convolutional layer, then calculate RUL. The RUL calculation includes global average pooling layer, dense layer and relu layer. The ith element of feature attained by global average pooling layer, and predicted RUL can be mathematically described as

$$p(i) = \frac{1}{H}\sum_{j=0}^{H} c_3^i(j) \tag{9}$$

$$RUL = \sigma(W_r\sigma(W_d p{+}b_d) + b_r) \tag{10}$$

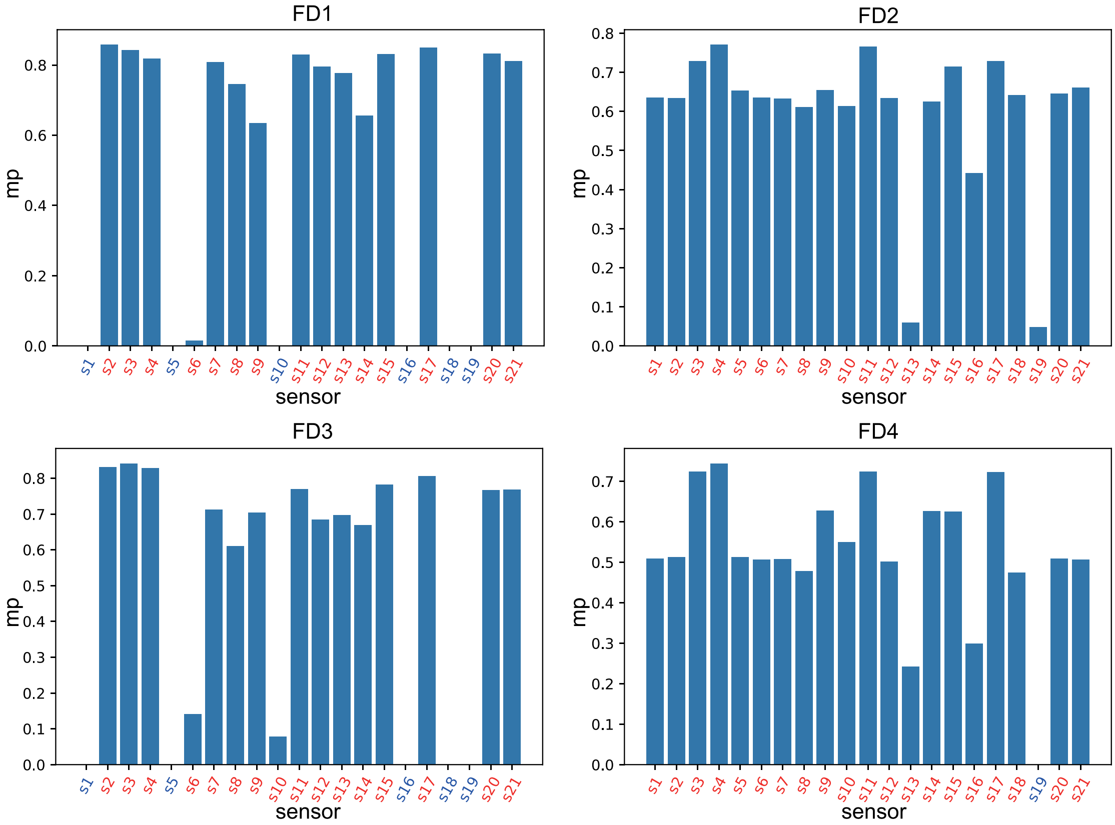


Fig. 3. Selection results by LM signal selection method (Red indicates the selected sensor).

Table 1
Details of C-MAPSS dataset.

Dataset	C-MAPSS			
	FD1	FD2	FD3	FD4
Number of engine for training	100	260	100	249
Number of engines for testing	100	259	100	248
Number of operating conditions	1	6	1	6
Number of fault modes	1	1	2	2
Number of training samples	17,731	48,819	21,820	57 522
Number of testing samples	100	259	100	248

where σ, H, and $c_3^i(j)$ indicate rectified linear unit, the length of the input feature, and the jth element of the ith channel of the c_3 attained by fully convolutional layer. $W_d \in \mathbb{R}^{64\times64}$, $W_r \in \mathbb{R}^{64\times1}$ are the weights of the first dense layer and the second dense layer. b_d and b_r are the biases of the first dense layer and the second dense layer.

3. Experiments

In this section, we will introduce the experimental setup, ablation experiments, and comparative experiments. The ablation experiment aims to prove the necessity and advanced performance of LM and TaNet. The comparative experiments were conducted to demonstrate the superiority of the prognostic method.

3.1. Experimental setup

3.1.1. C-MAPSS dataset

Our prognostic method was verified in the C-MAPSS dataset [27]. It is made up of data collected from the C-MAPSS turbofan engine, which was equipped with 21 sensors to collect data on different parts and locations. The data was divided into four subdatasets, including FD1, FD2, FD3, and FD4. Each subdataset was composed of a training set and a testing set. The engines of the Training set and testing set are the same type of engines, but their initial wear and manufacturing changes may be different. This is unknown to the users of the C-MAPSS dataset. In the training set, each multivariate time series data starts with the normal engine. After some cycle, the faults appear, and then the degree of failure continues to deepen. Finally, each multivariate time series data ends with a totally failed engine. Unlike the training set, the data of the testing set end up before the engine fails completely.

The data was divided into four subdatasets, including FD1, FD2, FD3, and FD4. Each subdataset was composed of a training set and a testing set. Table 1 lists the details of the C-MAPSS dataset. The data from FD2 and FD4 was collected from the most diverse operating conditions. More operating conditions may make the samples of the same RUL label value have more different characteristics, causing performance degradation for RUL prediction. Resistance of RUL prediction also derives from the diversity of fault modes. It is easy to analyze that FD4 is the most challenging task in the four subdatasets.

3.1.2. Data processing

As represented in Fig. 3, we selected those signals whose mp is greater than zero for each subdataset. The selected features is marked in red. To accelerate the speed of solving the optimal solution in gradient descent, each feature dimension of samples was normalized by standardscaler normalization method, which is given by

$$x_{norm}^i(j) = \frac{x^i(j) - \mu^i}{\sigma^i} \tag{11}$$

where $x_{norm}^i(j)$ denotes the jth data point of ith feature dimension after normalization, and $x^i(j)$ denotes the jth data point of ith feature dimension before normalization. μ^i is the mean of all sample in the ith feature dimension, and σ^i is the standard deviation of all sample in the ith feature dimension.

For RUL prediction, a component or equipment has been in a normal state for an extended period in its entire life cycle. Hence, Ramasso

Table 2
Parameter setting of training process.

Hyper-parameters	Description	Option
Batch size	The samples number of one back propagations	1024
Optimizer	Algorithm for minimizing loss value	Adam
Training epochs	The back propagations number of each samples	2000
EarlyStopping	The epochs with no improvement on training loss	50
$Patience_{ReduceLROnPlateau}$	The epochs with no improvement on testing loss	20
$Factor_{ReduceLROnPlateau}$	Factor used to reduce the learning rate	0.5
$LR_{ReduceLROnPlateau}$	Minimum learning rate	0.0001

Table 3
Parameter number ratio of attention mechanisms.

Attention mechanism	A2 [16]	A1 [15,23]	A0 [20]	TaNet
Ratio	6.46	0.62	1	1

proposed a piece-wise linear degradation model For the C-MAPSS turbofan engine dataset [28]. This model set the sample of turbofan engine at an early age to a constant label. In most cases, the sliding average value of signals began to change rapidly when the remaining useful life value was around 115, which indicated that the health status of the device began to change at this time, so it was set to 115 in our experiments. The model assumed that the engine works typically in the early stages, and its life span is constant. When a failure occurs, life begins to degrade linearly. Eventually, the engine fails. Analyzed from the general law of equipment or component degradation, the hypothesis is reasonable and practical.

Multiple time step series contain more information than single time step data. More than that, multiple time step series contains the change trend information of signal data, which is conducive to RUL prediction. Hence, we chose the longest time step for each subdataset. The time window sizes of FD1, FD2, FD3, and FD4 were set to 31, 21, 38, and 19.

3.1.3. Performance metrics

To improve the comprehensiveness of model evaluation, we employed two metrics: root mean square error (RMSE) and Scoring function (S).

RMSE is a popular performance metric for prediction task. It is given by

$$RMSE = \sqrt{\frac{1}{Q}\sum_{i=1}^{Q}\left(y_i^{pred} - y_i\right)^2} \tag{12}$$

where Q is the number of samples, and y_i^{pred} and y_i refers to the estimated RUL value and real RUL value of the ith sample, respectively.

Unlike RMSE, S is a performance metrics designed for RUL prediction tasks. It can be mathematically expressed as

$$S = \sum_{i=1}^{Q} s_i \tag{13}$$

$$s_i = \begin{cases} e^{\frac{y_i - y_i^{pred}}{13}} - 1 & y_i^{pred} - y_i < 0 \\ e^{\frac{y_i^{pred} - y_i}{10}} - 1 & y_i^{pred} - y_i \geq 0 \end{cases} \tag{14}$$

It can be analyzed from the Eq. (14) that when the estimated RUL value is higher than the real RUL value, the score function penalty is more severe. Hence, the score function is biased towards choosing a model that predicts early failure rather than a model that predicts delayed failure.

3.1.4. Prognostic procedure

Our prognostic procedure contains data pre-processing, training, and testing. As demonstrated in Fig. 4, our model training adopted EarlyStopping and ReduceLROnPlateau strategies. When validation loss ceases decreasing, the EarlyStopping terminates model training to avoid the over-fitting problem. Another strategy enables the learning rate to be continuously reduced during the training process. These two strategies can improve the performance of the model. For the best performance of our model, the hyper-parameters setting of the training process is listed in Table 2.

3.2. Ablation experiments for LM and TaNet

To prove the necessity and advanced performance of LM and TaNet, we conducted a chain of ablation experiments. Results demonstrated that our proposed LM and TaNet can improve model performance significantly compared to comparative methods, especially when combined.

3.2.1. Description

Ablation experiments adopted eight compared methods. All methods have a different combination in signal selection method and attention mechanism. The eight combinations contain LM+TaNet, F0+TaNet, F1+TaNet, F1 only, LM+A0, LM only, LM+A1, and LM+A2. The F0 and F1 refer to the signal selection method selecting all signals [10] and currently the most popular signal selection method of the C-MAPSS dataset, namely the monotonic discriminant method [4,15–18]. A0, A1, and A2 refer to the SeNet attention mechanism [20], the attention mechanism in [15,23], and the attention mechanism in [16]. The parameter number ratio of attention mechanisms in FD1 is shown in Table 3. Notably, the number of parameters of the TaNet we proposed is far less than that of A2.

To guarantee the reliability of experiment results, we conducted 10 trials on each subdataset for all methods, and then average values are presented. One thing to note, we found a tiny percentage of experiments where training set loss ceases change with a highly massive value in the first epoch of training. The model falls into a local optimum in the first epoch. This situation can be discovered in the first epoch of model training rather than during model testing. To guarantee the rationality of the results, we discarded the results obtained by this type of experiment.

To comprehensively evaluate the performance of the model on different datasets, we employed $RMSE_{sum}$ and S_{sum} to comprehensively and accurately evaluate the performance of the model in failure prediction, early failure prediction, and delayed failure prediction. Table 4 presents the results of ablation experiments. LM+TaNet and LM+A2 achieve the best performance in almost all indicators. Furthermore, LM+TaNet attained the lowest S_{sum}.

3.2.2. Analysis for LM

To quantify the performance difference between combinations, we adopted Wilcoxon signed-rank test [29], which judges whether two paired samples come from the same distribution. When the p-value calculated by Wilcoxon signed-rank test is less than 0.05, the hypothesis that two paired samples come from the same distribution should be rejected. Each sample is composed of $RMSE_{fd1}$, S_{fd1}, $RMSE_{fd2}$, S_{fd2}, $RMSE_{fd3}$, S_{fd3}, $RMSE_{fd4}$, S_{fd4}, $RMSE_{sum}$, S_{sum}.

As seen from Fig. 5 and Table 4, the p-values of (LM+TaNet)-(F1+TaNet) is 0.0039, which is far less than 0.05. That means that the paired samples have conspicuous differences. Moreover, all indicators of (LM+TaNet) are better than those of (F1+TaNet). The two powerful pieces of evidence prove that our proposed LM was significantly superior to the most popular signal selection method F1 of the C-MAPSS dataset. Our proposed LM selected signals by the mapping ability of the signal, so it can more accurately determine whether the signal data contains RUL information. That undoubtedly intensifies the performance of RUL prediction tasks. Meanwhile, the p-values of (LM+TaNet)-(F0+TaNet) is 0.21 and the $RMSE_{sum}$ and S_{sum} of (LM+TaNet) are slightly better than those of (F0+TaNet). That means that the extra signals selected by F0 did not contribute to model

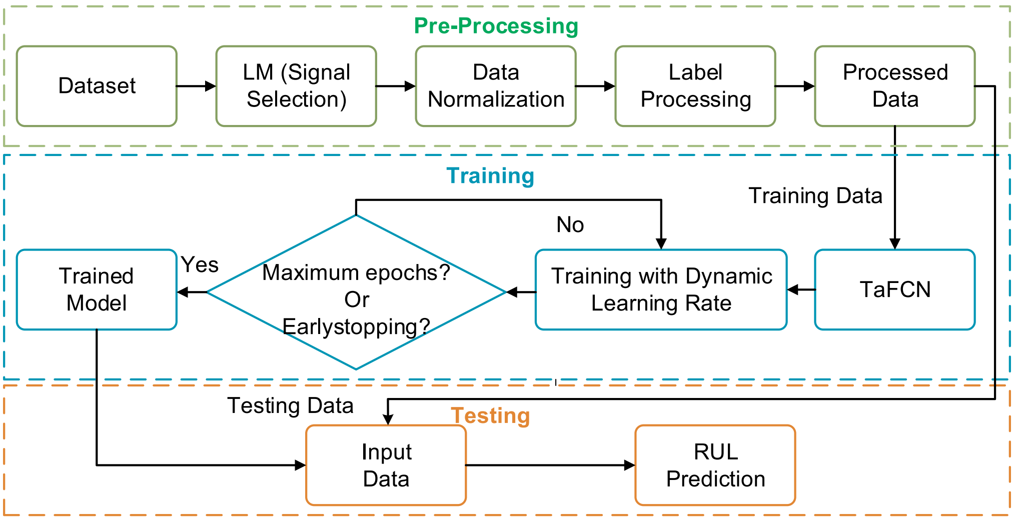
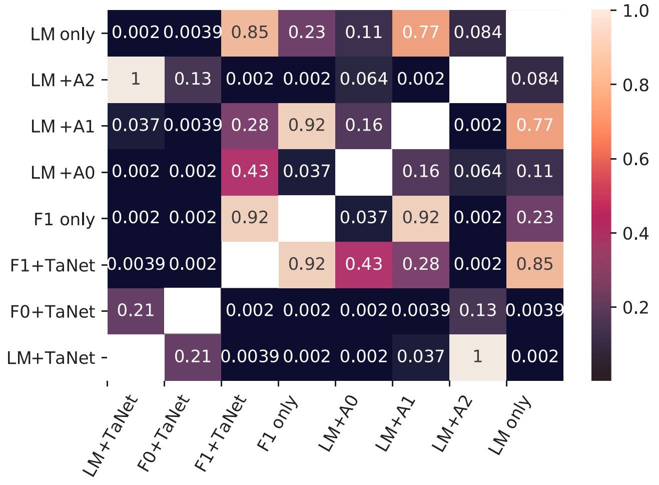


Fig. 4. Flowchart of prognostic procedure.



	LM+TaNet	F0+TaNet	F1+TaNet	F1 only	LM+A0	LM+A1	LM+A2	LM only
LM only	0.002	0.0039	0.85	0.23	0.11	0.77	0.084	
LM +A2	1	0.13	0.002	0.002	0.064	0.002		0.084
LM +A1	0.037	0.0039	0.28	0.92	0.16		0.002	0.77
LM +A0	0.002	0.002	0.43	0.037		0.16	0.064	0.11
F1 only	0.002	0.002	0.92		0.037	0.92	0.002	0.23
F1+TaNet	0.0039	0.002		0.92	0.43	0.28	0.002	0.85
F0+TaNet	0.21		0.002	0.002	0.002	0.0039	0.13	0.0039
LM+TaNet		0.21	0.0039	0.002	0.002	0.037	1	0.002

Fig. 5. Wilcoxon signed rank test comparison of eight combinations.

performance, and F0 is likely to degrade model performance. Hence, our proposed LM can choose effective signals accurately instead of blindly selecting more signals to improve the performance of RUL prediction tasks.

Noticeably, the *p*-value of (F1+TaNet)-(F1 only) is greater than 0.05, suggesting our proposed TaNet did not work when combined with the monotonic discriminant method. To explore the reasons why it happens, we conducted the following experiments. Unlike in other subdatasets where only one or almost all monotonic-discriminant-unselected sensors are selected, in FD3, LM selects two monotonic-discriminant-unselected sensors. This is a unique and interesting case that can help us analyze the effectiveness of LM. Therefore, the experiments are conducted on FD3.

For F1+TaNet and LM+TaNet, we performed the statistical analysis for the intermediate output value of the testing set. As demonstrated in Fig. 6, the deviations of the average attention values in F1+TaNet relative to 0.4 are mostly less than 0.1. It is because TaNet considered the importance of different signals to be similar most of the time. The signals attained by the monotonic discriminant method whose data value all changes monotonously. This means these signals likely to have similar characteristics. Hence, the behavior of TaNet is reasonable.

Compared with F1+TaNet, LM+TaNet made a big difference in attention distribution. LM selected two more signals than the monotonic discriminant method, which are S6 and S10. After adding two signals, the distribution of importance changes significantly. The average attention values of signals are concentrated around 0.3, 0.4, and 0.6, and the box plot of LM+TaNet has some extremely high attention values. These signals whose values are greater than 0.8 are likely to be highly helpful for RUL prediction, which demonstrates that the signal data of some sensors become more useful for the overall prediction task after adding S10 than before adding S10. This is because there may be synergies between different sensors in the RUL prediction task. Although the ability of S10 signal data alone to map RUL is weak, the effects of S9, S12, and S13 on the prediction task are significantly enhanced under the action of S10 signal data. This shows that S10 is essential for the RUL prediction task based on multiple sensor signal data. Significantly, the average attention value of S10 is at the highest level, which means TaNet considered that S10 added by LM is of great importance for the RUL prediction task. The high attention value of S10 comes from the average of all test samples, not a small number of samples. Furthermore, the effectiveness of TaNet has been proven by a series of ablation experiments and comparative experiments. Hence, this evidence is powerful. Smartly, the attention value attained by TaNet proves the effectiveness of LM.

3.2.3. Analysis for TaNet

As shown in Fig. 5, the *p*-value of (LM+TaNet)-(LM only), (LM+TaNet)-(LM+A0), and (LM+TaNet)-(LM+A1) is less than 0.05. Moreover, Table 4 present that the performance of (LM+TaNet) is better than that of LM only, LM+A0, and LM+A1. These two facts demonstrate that the performance of LM+TaNet is significantly prior to that of LM only, LM+A0, and LM+A1. Therefore, under the premise of our proposed LM, our proposed TaNet can significantly improve the network structure compared to not using the attention mechanism or using the A0 and A1 attention mechanism.

The poor performance of A0 and A1 is probable to be due to the inability to effectively characterize the feature samples and the inability to explore the correlation between features, respectively, and thus cannot focus on vital features. The *p*-value of (LM+TaNet)-(LM+A2) is 1, which means no significant difference between the paired samples. For the RMSE index, the performance of LM+TaNet is reduced by 0.72% relative to that of LM+A2. For the Sum_Score index, the performance of LM+TaNet is improved by 2.02% relative to that of LM+A2. These demonstrate that our proposed TaNet achieved slightly better performance than A2 in the RUL prediction task. Significantly, the computational cost of our proposed TaNet is only equivalent to 15.5% of the computational cost of A2. Hence, there is no doubt that our proposed TaNet is more suitable for RUL prediction tasks than A2.

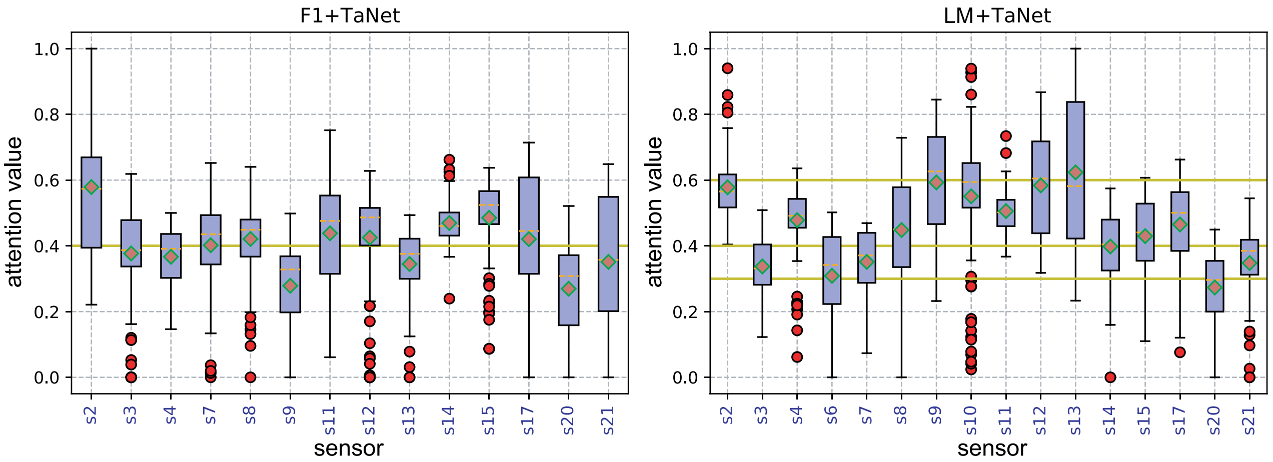


Fig. 6. Attention analysis of TaNet.

Table 4
Results of ablation experiment.

	LM+TaNet	F0+TaNet	F1+TaNet	F1 only	LM+A0	LM+A1	LM+A2	LM only
$RMSE_{fd1}$	13.99	13.93	13.98	14.75	14.50	13.84	**13.45**	14.63
S_{fd1}	336.45	329.66	345.85	376.41	365.46	330.50	**320.30**	384.19
$RMSE_{fd2}$	**17.06**	**17.06**	17.72	17.40	17.28	17.98	17.38	17.20
S_{fd2}	**1946.31**	**1946.31**	2623.05	2421.15	2207.19	2700.55	2366.39	2051.74
$RMSE_{fd3}$	12.01	11.92	12.76	13.40	12.78701	12.58	**11.74**	13.16
S_{fd3}	**251.26**	271.33	304.09	381.14	320.07	321.09	266.48	373.14
$RMSE_{fd4}$	**19.79**	19.99	20.32	20.51	20.00	20.24	19.83	19.97
S_{fd4}	3670.83	3944.89	4186.33	4057.90	4086.60	4332.00	**3379.3**	4362.44
$RMSE_{sum}$	62.85	62.91	64.77	66.07	64.56	64.65	**62.40**	64.96
S_{sum}	**6204.85**	6492.20	7459.32	7236.61	6979.33	7684.14	6332.47	7171.52

Table 5
Performance comparisons of the proposed prognostic method and the latest SOTA methods on the C-MAPSS dataset.

Method	RMSE/S					
	FD1	FD2	FD3	FD4	Sum	Win
MODBNE	15.04/334	25.05/5585	12.51/422	28.66/6558	81.26/12 899	0
DCNN	**12.61**/274	22.36/10 412	12.64/284	23.31/12 466	70.92/23 436	1
LSTM+FNN	16.14/338	24.49/4450	16.18/852	28.17/5550	84.98/11 190	0
SBRNN	13.58/**228**	19.59/2650	19.16/1727	22.15/**2901**	74.48/7506	2
Our method	13.99/336	**17.06/1946**	**12.01/251**	**19.79**/3671	**62.85/6204**	**7**

Table 6
Performance comparisons about error range.

Method	FD001	FD002	FD003	FD004
DLSTM	[−47, 56]	N/A	[−44, 38]	N/A
Our method	[−36.37, 35.05]	[−43.74, 57.48]	[−26.59, 31.10]	[−59.72, 65.09]

The ablation experiment proves that our proposed TaNet achieved better performance than the previous attention mechanisms at a small computational cost.

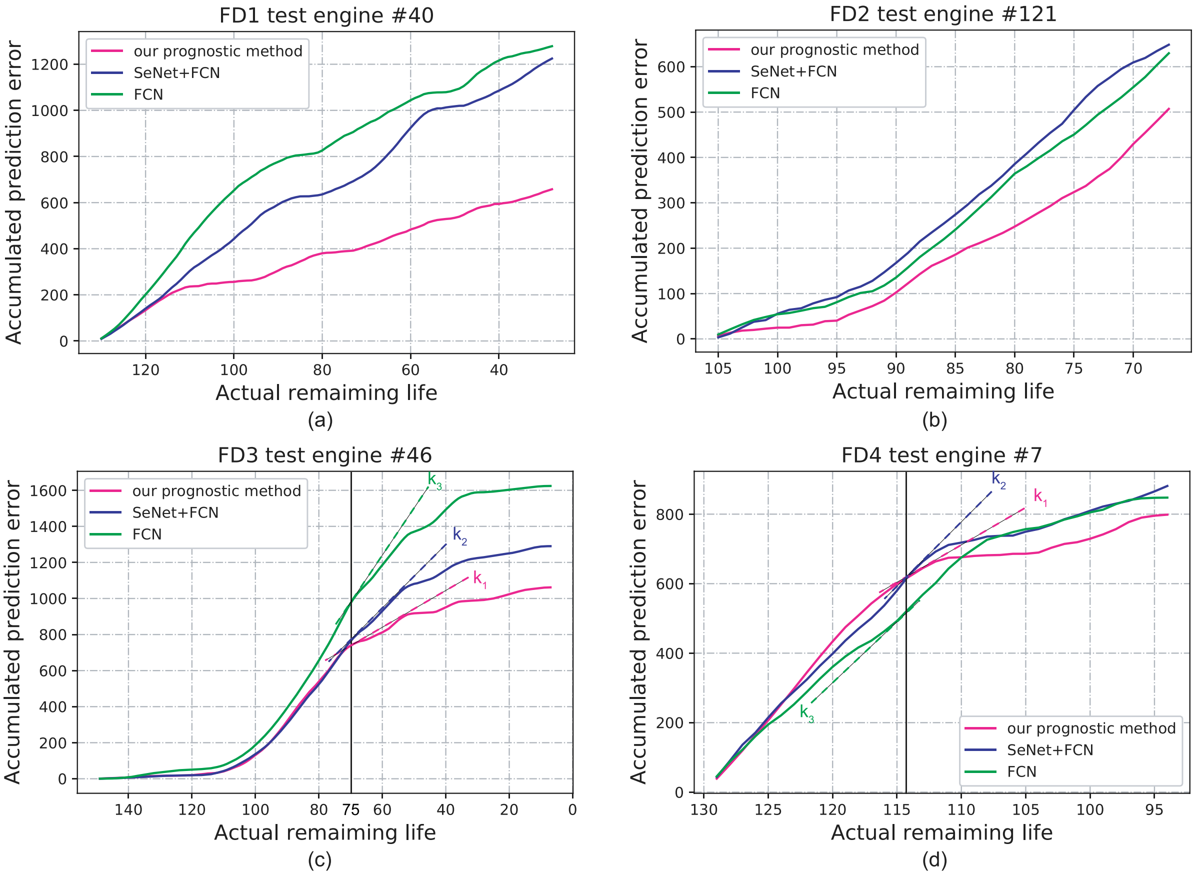
3.3. Comparing with SOTA methods

To demonstrate the superiority of our prognostic method, we compared it with the latest SOTA methods [4,17,19,30,31] on the C-MAPSS dataset. Table 5 summarizes comparison between our method with other methods.

Table 5 shows that our method achieved the best performance in all indicators of FD2 and FD3, and the best performance in the RMSE indicator of FD4. Subsequently, we calculated the sum of all indicators. The smaller the value, the better the performance of this method on the C-MPASS turbofan engine dataset. Not surprisingly, our method has the smallest sum values, 11.38% and 17.35% smaller than the second RMSE and S, respectively. Moreover, among the 10 indicators, our method became the winner in 7 indicators, far better than other methods. These indicate that our method has state-of-the-art capability in RUL prediction tasks.

We compared our method with only one method in Table 6 due to the highly limited paper reporting error range. For early RUL and late RUL, the range obtained by our method is far narrower than that obtained by DLSTM [19], which indicate that our method provided stable accuracy for RUL prediction tasks.

Fig. 7 shows the accumulated prediction error over RUL of our prognostic method and comparative methods. It refers to the sum of prediction errors from the initial RUL to the current RUL. Fig. 7(a) and (b) shows the accumulated prediction errors of our prognostic method from the initial RUL to the final RUL is less than that of other methods, and the difference value gradually expands with RUL. That shows that our method performs better in the full lifecycle RUL prediction of FD1 test engine#40 and FD2 test engine#121. Fig. 7(c) and (d) show that in the first half of the lifecycle, the accumulated prediction errors of other methods that of our prognostic method are comparable. However, our method performs better in the second half of the lifecycle and expands



(a)

(b)

(c)

(d)

Fig. 7. Accumulated prediction error over RUL.

its advantages with RUL. Take Fig. 7(c) as an example, RUL = 75 is the change point, where the slope of the curve $k_1 > k_2 > k_3$. The smoother the curve, the smaller the prediction error. After RUL = 75, our method dominate in accumulated prediction error. Moreover, because the curve slope of our method is smaller than that of other methods in the second half of the life cycle, our advantages in accumulated prediction error continue to expand with RUL.

By the comparative experiments, it was proven that our prognostic method has stable and state-of-the-art performance in RUL prediction tasks.

4. Interpretability

Deep learning provides extraordinary predictive capabilities. However, as a black-box model, it is difficult to be understood and trusted. Hence, users urgently need to understand the prediction behavior of deep models, how data features affect prediction, and the sensitivity to different data features.

In this section, we will analyze the logic reasoning of our prognostic method. The time window size of FD3 is the longest timestep in all subdatasets. Choosing FD3 can help us analyze the importance of the activation at temporal location t leading to the prediction results at the largest time scales. Therefore, the experiments of interpretability are conducted on FD3. The structure of Class Activation Mapping (CAM) [32] is designed for image classification tasks. Hence, to make it suitable for prediction methods, so we modified it to some extent. We proposed the calculation method of convolutional channel weight for multi-dense layer prediction structure, which can reversely calculate the weight of each channel of the last convolutional layer in the neural network to construct the temporal importance vector when predicting RUL. The definition of modified CAM is as follows:

$$W = W_d W_r \tag{15}$$

$$L_{CAM} = \sum_{i=1}^{n} W^i A^i \tag{16}$$

$$L^t_{CAM} = \frac{L^t_{CAM} - \min(L_{CAM})}{\max(L_{CAM}) - \min(L_{CAM})} \tag{17}$$

$$L^t = \frac{\exp(L^t_{CAM})}{\sum_{t=1}^{l} \exp(L^t_{CAM})} \tag{18}$$

where W^i, A^i, n, l, L_{CAM}, L^t_{CAM}, and L^t refers to the ith element of W, the ith feature vector attained by the last convolutional layer, the length of an input feature vector and the filter size of last convolutional layer, the original temporal importance vector, the value of the original temporal importance vector at temporal location t, and the importance of the activation at temporal location t leading to the prediction result, respectively. $W_d \in \mathbb{R}^{64\times64}$ and $W_r \in \mathbb{R}^{64\times1}$ are the weights of the first dense layer and the second dense layer.

In Eq. (15), the weights of the first dense layer are multiplied by the weights of the second dense layer matrix to obtain the channel weight of the last convolutional layer. It should be noted here that the hth row vector in $W_d \in \mathbb{R}^{64\times64}$ is the weights between all neurons of the second dense layer and the hth neurons of the first dense layer. In this way, the reverse calculation of the weight of the hth channel can be realized. In Eq. (16), the weight of each channel is multiplied by the corresponding feature vectors attained by the last convolutional layer to obtain the temporal importance vector of the sample. Then, a Min–Max scaling was conducted in Eq. (17), all values of temporal importance vector L_{CAM} are converted to [0,1]. Finally, in Eq. (18), we conducted a softmax activation on the scaled L_{CAM}, which makes the sum of all values of the temporal importance vector equal to 1.

Sample point value is denoted as $v_t = \sum_{k=1}^{num} v^k_t$, where k, n, t, and v^k_t are the index of the sensors selected by LM, numbers of sensors selected

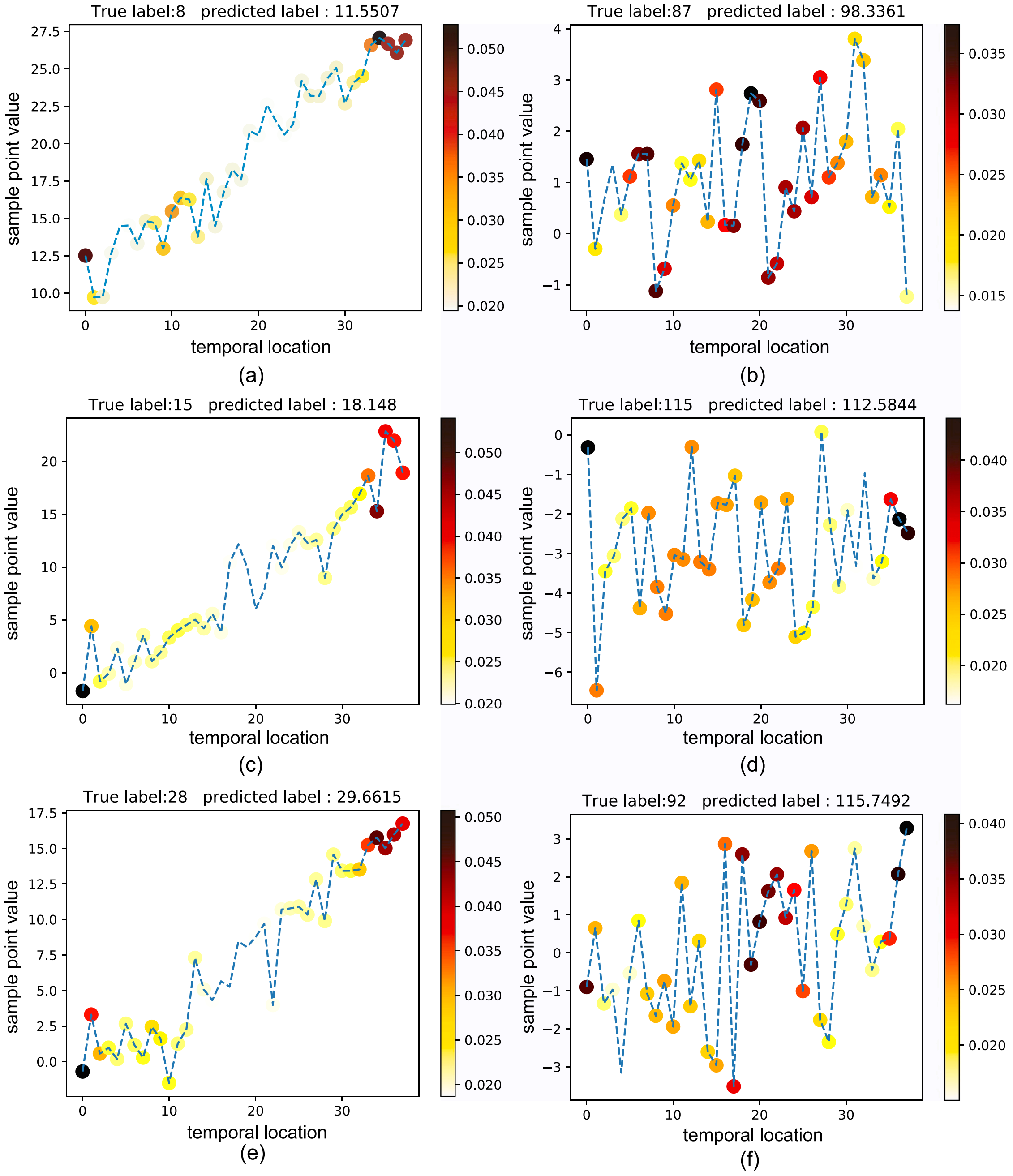


Fig. 8. Interpretability analysis for the prognostic method.

by LM, temporal location t of sample timestep, and the signal value of the kth sensor of the sample at temporal location t, respectively. Fig. 8 shows two kinds of typical figures obtained from rectified CAM. The value on Fig. 8(b), (d), and (f) oscillates up and down, and the value on Fig. 8(a), (c), and (e) rises monotonously. The darker the color of a data point is, the greater the data point contributed to prediction results.

Fig. 8 indicates the first point always plays the most crucial role in all points. The first point in the time series data is the benchmark for all other points, so there is no doubt that it is of great importance. The left pictures in Fig. 8 show that when the data points tend to rise monotonously, the color of the data points at both ends is darker, and the color of the middle data point is almost invisible. That shows that when the data points are rising monotonically, our model did not care about the middle data points of the sample but relies on the values of the data points at both ends to predict RUL. According to the upward trend of the sample, our proposed model utilized the information of the data points at both ends to characterize the sample information. That reduced the interference of useless information in the intermediate point, so it is an intelligent approach. The right pictures in Fig. 8 shows that when the data point oscillates up and down, the color of the middle data point is darker. That shows that our model paid more attention to the information at the intermediate points to predict RUL when the data points oscillate up and down. Because the data points oscillate up and down, our model cannot obtain all sample information through the data points at both ends, so it turns to the middle data point to obtain more information about RUL. It is obvious that vital information is stored at both ends or in the middle. Hence, to represent sample information accurately and efficiently, we set N to 3. In this way, the all information of a sample is converted into 3 values to characterize the trend information of RUL monitoring signals. The distribution of vital information in the sample demonstrates that the setting of N and the proposed representation method in TaNet are reasonable.

We developed a modified CAM method which is a general interpretability analysis method for convolutional neural network prediction models. The experimental results show that when our prognostic method processes different samples, by identifying the pattern of the samples, it can intelligently give different degrees of attention to the sample points and extract meaningful information for the RUL prediction tasks.

5. Conclusion

This paper proposed a novel prognostic method for multivariate RUL estimation tasks, composed of the LM signal selection method and the TaFCN end-to-end RUL prediction framework. In TaFCN, the proposed TaNet attention mechanism is specially designed by the characteristics of RUL monitoring signals. Furthermore, an interpretability analysis method was developed for RUL prediction tasks for the first time. With the prognostic method, we can accurately select useful signals based on the ability of signal data to map RUL and assign optimal attention levels to the signals at a small computational cost. Based on those, TaFCN achieved a high-precision RUL prediction.

To demonstrate the necessity and advanced performance of LM and TaNet, we conducted a series of ablation experiments of the latest signal selection methods and attention mechanisms. The results attained by Wilcoxon signed rank test demonstrate that our proposed LM and TaNet both can significantly improve the model performance compared to the comparative methods, especially when the two are combined. Smartly, the effectiveness of LM was proven by the attention levels assigned by TaNet in the ablation experiments. Additionally, to prove the superiority of our prognostic method, we compared it with four state-of-the-art methods. The comparison results indicate that our prognostic method outperforms them. That means the proposed prognostic method has state-of-the-art performance for RUL prediction tasks. Besides, we developed a modified CAM to obtain the logical reasoning process of the prognostic method. Results indicate that the method intelligently employed data points at the two ends or in the middle to extract vital information for RUL predict tasks..

In this research, we applied TaNet to the feature dimension. Inspired by the different effects of multiple time steps on the prediction results, our future research will focus on the combination of TaNet and gate structure in time and feature dimensions.

CRediT authorship contribution statement

Linchuan Fan: Writing – original draft, Visualization, Software, Resources, Methodology, Formal analysis, Conceptualization. **Yi Chai:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Conceptualization. **Xiaolong Chen:** Writing – review & editing, Validation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by National Natural Science Foundation of China (NSFC) grant no. U2034209.

References

[1] Lei Y, Li N, Guo L, Li N, Yan T, Lin J. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mech Syst Signal Process 2018;104:799–834.

[2] Liao L, Köttig F. Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. IEEE Trans Reliab 2014;63(1):191–207.

[3] Wang B, Lei Y, Li N, Li N. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. IEEE Trans Reliab 2018;69(1):401–12.

[4] Li X, Ding Q, Sun J-Q. Remaining useful life estimation in prognostics using deep convolution neural networks. Reliab Eng Syst Saf 2018;172:1–11.

[5] Yu W, Shao Y, Xu J, Mechefske C. An adaptive and generalized Wiener process model with a recursive filtering algorithm for remaining useful life estimation. Reliab Eng Syst Saf 2022;217:108099.

[6] Duan F, Wang G. Bayesian analysis for the transformed exponential dispersion process with random effects. Reliab Eng Syst Saf 2022;217:108104.

[7] Li X, Zhang W, Ding Q. Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. Reliab Eng Syst Saf 2019;182:208–18.

[8] Yang B, Liu R, Zio E. Remaining useful life prediction based on a double-convolutional neural network architecture. IEEE Trans Ind Electron 2019;66(12):9521–30.

[9] Cao Y, Ding Y, Jia M, Tian R. A novel temporal convolutional network with residual self-attention mechanism for remaining useful life prediction of rolling bearings. Reliab Eng Syst Saf 2021;215:107813.

[10] Peng C, Chen Y, Chen Q, Tang Z, Li L, Gui W. A remaining useful life prognosis of turbofan engine using temporal and spatial feature fusion. Sensors 2021;21(2):418.

[11] Zhuang J, Jia M, Ding Y, Ding P. Temporal convolution-based transferable cross-domain adaptation approach for remaining useful life estimation under variable failure behaviors. Reliab Eng Syst Saf 2021;216:107946.

[12] Wang Z, Yan W, Oates T. Time series classification from scratch with deep neural networks: A strong baseline. In: 2017 international joint conference on neural networks (IJCNN). IEEE; 2017, p. 1578–85.

[13] Karim F, Majumdar S, Darabi H, Chen S. LSTM fully convolutional networks for time series classification. IEEE Access 2017;6:1662–9.

[14] Bagnall A, Lines J, Bostrom A, Large J, Keogh E. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. Data Min Knowl Discov 2017;31(3):606–60.

[15] Chen Z, Wu M, Zhao R, Guretno F, Yan R, Li X. Machine remaining useful life prediction via an attention-based deep learning approach. IEEE Trans Ind Electron 2020;68(3):2521–31.

[16] Liu H, Liu Z, Jia W, Lin X. Remaining useful life prediction using a novel feature-attention-based end-to-end approach. IEEE Trans Ind Inf 2020;17(2):1197–207.

[17] Yu W, Kim IY, Mechefske C. An improved similarity-based prognostic algorithm for RUL estimation using an RNN autoencoder scheme. Reliab Eng Syst Saf 2020;199:106926.

[18] Shi Z, Chehade A. A dual-LSTM framework combining change point detection and remaining useful life prediction. Reliab Eng Syst Saf 2021;205:107257.

[19] Wu J, Hu K, Cheng Y, Zhu H, Shao X, Wang Y. Data-driven remaining useful life prediction via multiple sensor signals and deep long short-term memory neural network. ISA Trans 2020;97:241–50.

[20] Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-excitation networks. IEEE Trans Pattern Anal Mach Intell 2020;42(8):2011–23.

[21] Artacho B, Savakis A. UniPose+: A unified framework for 2D and 3D human pose estimation in images and videos. IEEE Trans Pattern Anal Mach Intell 2021;1.

[22] Zhao Z, Li Q, Zhang Z, Cummins N, Wang H, Tao J, Schuller BW. Combining a parallel 2d cnn with a self-attention dilated residual network for ctc-based discrete speech emotion recognition. Neural Netw 2021;141:52–60.

[23] Liu C, Zhang L, Niu J, Yao R, Wu C. Intelligent prognostics of machining tools based on adaptive variational mode decomposition and deep learning method with attention mechanism. Neurocomputing 2020;417:239–54.

[24] Xiang S, Qin Y, Zhu C, Wang Y, Chen H. LSTM networks based on attention ordered neurons for gear remaining life prediction. ISA Trans 2020;106:343–54.

[25] Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. In: Icml. 2010.

[26] Fawaz HI, Forestier G, Weber J, Idoumghar L, Muller P-A. Deep learning for time series classification: a review. Data Min Knowl Discov 2019;33(4):917–63.

[27] Saxena A, Goebel K. Turbofan engine degradation simulation data set. In: NASA Ames prognostics data repository. Moffett Field, CA: NASA Ames Research Center; 2008.

[28] Ramasso E. Investigating computational geometry for failure prognostics. Int J Progn Health Manage 2014;005(1):1–18.

[29] Wilcoxon F. Individual comparisons by ranking methods. In: Breakthroughs in statistics. Springer; 1992, p. 196–202.

[30] Zheng S, Ristovski K, Farahat A, Gupta C. Long short-term memory network for remaining useful life estimation. In: 2017 IEEE international conference on prognostics and health management (ICPHM). IEEE; 2017, p. 88–95.

[31] Zhang C, Lim P, Qin AK, Tan KC. Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics. IEEE Trans Neural Netw Learn Syst 2017;28(10):2306–18.

[32] Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2016, p. 2921–9.