IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 10, 2018, accepted February 6, 2018, date of publication February 12, 2018, date of current version March 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2804930

Prediction of Bearing Remaining Useful Life With Deep Convolution Neural Network

LEI REN[1, 2, 3], (Member, IEEE), YAQIANG SUN[1], HAO WANG[4], (Member, IEEE), AND LIN ZHANG[1, 2, 3], (Senior Member, IEEE)

[1]School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China
[2]Engineering Research Center of Complex Product Advanced Manufacturing System, Ministry of Education, Beijing 100191, China
[3]Beijing Advanced Innovation Center for Big Data-based Precision Medicine, Beihang University, Beijing 100191, China
[4]Department of ICT and Natural Sciences, Norwegian University of Science and Technology, 6025 Aalesund, Norway

This work was supported in part by the National Science Foundation of China under Grant 61572057 and in part by the National High Technology Research and Development Program of China under Grant 2015AA042101.

ABSTRACT Cyber-physical-social system (CPSS) has drawn tremendous attention in industrial applications such as industrial Internet of Things (IIoT). As the fundamental component of IIoT, bearings play an increasingly important role in CPSS for IIoT. Better understanding of bearing working conditions and degradation patterns so as to more accurately predict the remaining useful life (RUL), becomes an urgent demand for industrial prognostics in IIoT. The data-driven approach has indicated good potential, but the prediction accuracy is still not satisfactory. This paper proposes a new method for the prediction of bearing RUL based on deep convolution neural network (CNN). A new feature extraction method is presented to obtain the eigenvector, named the spectrum-principal-energy-vector. The eigenvector is suitable for deep CNN. In the prediction phase, we propose a smoothing method to deal with the discontinuity problem found in the prediction results. To the best of our knowledge, we are the first to propose such a smoothing method for bearing RUL prediction. Experiments show that our method can significantly improve the prediction accuracy of bearing RUL.

INDEX TERMS Cyber-physical-social system, industrial big data, deep learning, RUL prediction, deep convolution neural network.

I. INTRODUCTION

Cyber-Physical-Social System (CPSS) is considered as a newly emerged paradigm encompassing the cyber world, physical world and social world [1]. The Internet of Things (IoT), bridging the physical world and cyber world, has become an essential part of CPSS, providing the support for sensing, monitoring, and interpreting the environment. As a typical application of IoT in industry, Industrial Internet of Things (IIoT) has drawn much attention recently and provides new research opportunities for CPSS in industry. As the fundamental component of most industrial rotational equipment, bearings play an increasingly important role in IIoT. The health conditions of bearings have major impacts on the reliability and the operation accuracy of CPSS for IIOT: bearing failures could seriously reduce the production precision and even result in the equipment failure. Due to many complex factors such as material features and working environment, same types of bearings by the same manufacturer, working in same type of equipment, could have very different life expectancy and degradation pattern [2]. Many different factors have different impacts on the *remaining useful life* (RUL) of bearings [3]. Unfortunately, many of these factors cannot be quantitatively measured and analyzed, which makes the prediction of bearing RUL still a significantly challenging task [4]. How to improve the accuracy in the prediction of bearing RUL becomes an urgent problem and has attracted increasing amount of attention among prognostics researchers.

Recently, most proposed methods for the prediction of bearings RUL mainly fall into two categories: model-based approach and data-driven approach [5]. We observe that methods in the model-based approach are difficult to further improve the prediction accuracy (detailed discussions in Section II). On the other hand, in the data-driven approach, increasing amount of relevant data, such as temperature, load, speed, and bearing vibration amplitude, are being captured while the bearing is working [6]. The collected data can then be analyzed for the prediction of RUL. The core idea of data-driven approach is to analyze the current working condition of the bearing, with an attempt to find the relationship between the operating status and RUL expectancy [7]. The data on the operating status can effectively reflect the degradation of the bearing due to material defects and other factors. Therefore, in the data-driven approach, we can use these data reflecting degradation instead of directly quantifying the material defects and many other complicated factors. In this way, the data-driven approach becomes a more promising approach to the RUL prediction challenge, supported by the good results of recent proposed data-driven methods.

Due to the large amount of data, high data dimension, high interference noise and complicated mapping relationship, the traditional signal processing method and even the traditional machine learning method cannot capture implicit relations between different features in the bearing vibration data set. So, the prediction accuracy is not high. Due to the development of the deep learning technology [8], the ability to analyze complex data is greatly improved, and the deep learning method is increasingly used in prediction problems [9]. Deep learning technology can dig out the deeper level of information [8], so as to prominently improve the prediction accuracy. *Convolution Neural Network* (CNN) [10] plays an increasingly important role in deep leaning. Compared with deep neural network, CNN can use fewer parameters to achieve the same functionality or precision. Therefore, CNN is a good fit for high dimensional data, which is the most outstanding feature of bearing RUL data.

In this paper, we use the deep CNN model to predict the RUL of bearing. The experimental results showed that, in the bearing RUL prediction problem, our method can get better prediction accuracy. The contributions can be summarized as follows:

(1) We present a new feature extraction method, namely Spectrum-Principal-Energy-Vector, to obtain the eigenvector. This eigenvector can represent the decay of bearing vibration signal with the use time and is suitable for the structure of convolution neural network.

(2) Based on the characteristics of convolutional neural network, we propose a new prediction framework of the remaining life of bearing.

(3) We propose a post-smoothing method to address the discontinuity problem in prediction results. To the best of our knowledge, we are the first to propose such a smoothing method for the prediction of bearing RUL.

(4) We have conducted a comprehensive set of experiments with different feature extraction methods and machine learning prediction models for the prediction of RUL, and performed thorough comparisons. The experiment results show that our new method can significantly improve the prediction accuracy.

II. RELATED WORK

In data-driven bearing RUL prediction approach, the disparities between the different methods are mainly in two aspects: 1) feature vector, 2) prediction model.

The feature vector contains three types: time domain features, frequency domain features, time-frequency domain features [11]. The time domain features can characterize the degradation of the bearing. The full life vibration signal can be seen clearly. With the degradation of the bearing, the amplitude is gradually increasing. However, the time domain features are slow and fluctuating, so the life prediction based on time domain features is often not good. Frequency domain features have obvious advantages. Frequency domain transformation can change complex differential relationship into a linear relationship, reducing system complexity. At the same time, the frequency domain transformation can make the signal and noise separation, which can better suppress the noise information. However, since the original signal is transformed into frequency domain, the obtained frequency spectral information still has a high dimension. So most of the methods extract the feature of the spectrum, reducing the frequency domain feature dimension. Most of the frequency domain information is lost by the feature extraction of the spectrum. And this will reduce the prediction accuracy of the bearing's RUL prediction. The time-frequency domain feature is obtained by wavelet decomposition. Similar to the frequency domain feature processing method, the result of wavelet decomposition is still to be extracted, and the high-dimensional information is reduced to low-dimensional information.

These feature processing methods are mainly based on the traditional machine learning algorithm, and these methods meet the characteristics of the shallow machine learning model. If the feature dimension is high, the traditional machine learning algorithm, such as support vector machine [12], will show poor generalization ability. Thus, lower-dimensional features are suitable for traditional machine learning methods. The deep learning model has a good ability to deal with high-dimensional features. But with the traditional low-dimensional features, deep learning model can not improve the prediction accuracy. Because the traditional low-dimensional features loss too much information, and the low-dimensional features become the bottleneck of the prediction accuracy. Therefore, how to deal with the signal to ensure that information loss as small as possible is a key issue to improve the prediction accuracy. In this paper, a new feature construction method is proposed for the deep CNN.

In this paper, the method, transforming the original signal into the input features of the deep CNN, is proposed. The features suitable for the deep CNN is obtained, which is named as the Spectrum-Principal-Energy-Vector. After obtaining the Spectrum-Principal-Energy-Vector, it is input to the deep CNN. The deep CNN analyzes the input data and obtains a series of eigenvectors. Afterwards, the deep neural network model is used for regression prediction to obtain the RUL of the bearing.

III. METHODOLOGY

The framework for bearing RUL prediction is shown in the figure 1.

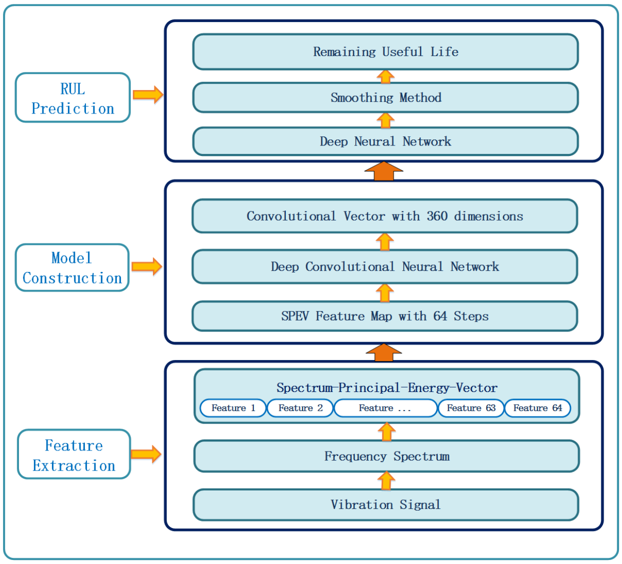


FIGURE 1. Bearings RUL Prediction Framework.

In the stage of feature extraction, begin with the original discrete vibration signal, the vibration signal is first subjected to Fast Fourier Transform (FFT) to produce the discrete frequency spectrum. The spectral dimension and the original vibration signal are 2560 dimensions. After the high-dimensional spectrum is obtained, the spectrum is divided into 64 blocks. In the frequency band in each block, the maximum amplitude in the frequency band is selected as the eigenvalue of the frequency band. The 64-dimensional Spectrum-Principal-Energy-Vector is obtained. In the stage of CNN, for each sampling time point, taken 63 samples from the front 63 sampling time points, the 64 Spectrum-Principal-Energy-Vectors will be combined into a feature map, and this feature map has the shape of (64*64). The feature map will be put into the deep CNN, get 360-dimension vector. Finally, in RUL Predict stage, using the 360-dimensional vector as input, the deep neural network will get the final prediction result.

A. FEATURE EXTRACTION

The original signal is the vibration signal of running bearing in many models [14], [15]. The traditional signal features have a large loss of information, such as the classic time domain features, the frequency domain features, and the time-frequency domain features. To ensure the integrity of information, it is important to mine information from the original data as far as possible. However, if using the 2560-dimensional vibration signal as a model input, it will lead to a large network structure. If the time step effect was considered, the input feature even reaches the million, and the huge network will lead to training difficulties, over-fitting and other shortcomings. In addition, the vibration signal contains a lot of noise information, and the noise information will seriously affect the prediction accuracy of the model. Because the vibration signal is difficult to separate the noise or useful information, so the traditional filtering method may cause the loss of key signal. But without de-noising, it will make the forecast network spend huge capacity to resolute noise, and it will have some impact of prediction accuracy. Time domain information is not suitable for forecasting model input directly. Taking the lack of time domain features into account, this paper uses frequency domain information for life prediction.

The vibration signal is subjected to discrete Fourier transform to obtain the spectrum of the signal. The spectrum has 2560 dimensions and so as to the vibration signal. In order to improve the prediction accuracy, this paper increases the number of prediction steps. The features of k time points, $t-k+1, t-k+2, \ldots, t$, are predicted as the total feature when the RUL at time t is predicted. The time step cannot be too long. In principle, it should be less than 100, because the total life of some bearings is between 100 to 200, too long a time-step will lead to the scope of application limit. The step in this paper is 64. But the structure of 2560 * 64 is clearly not suitable for deep CNN, because it will lead to a larger network, and it will incur difficulties in the subsequent parameter settings. Too large input shape will lead to a huge training model. In addition, in the 2560-dimensional spectrum information, there are many information have low correlation with RUL. In the frequency spectrum, the higher the amplitude is, the higher energy distribute. Therefore, this paper proposes the Spectrum-Principal-Energy-Vector to optimize the 2560-dimensional frequency spectrum.

The calculation of the Spectrum-Principal-Energy-Vector is as follows.

$x(i)$ for $i = 1, 2, \ldots, n$ is vibration signal. FFT is used to obtain the spectral sequence, $z(t)$ for $t = 1, 2, \ldots, n$.

$$z(t) = \sum_{n=0}^{N-1} A(n) W_N^{nk}, k = 0, \ldots, N-1 \tag{1}$$

$$W_N = e^{(-j2\pi/N)} \tag{2}$$

Performs a modulo operation on $z(t)$ to obtain a spectral sequence $s(j)$ for $j = 1, 2, \ldots, n$. And the SPEV index, $X_{SPEV}(k)$, will be calculated as:

$$X_{SPEV}(k) = max\{s(64*k-64), s(64*k-63), \ldots, s(64*k-1)\} \tag{3}$$

Where, $k = 1, 2, \ldots, K$, and $K = 64$ in this paper. $s(j)$ is equally divided into 64 segments. X_{SPEV} is a 64-dimensional column vector.

After the 64-dimensional Spectrum-Principal-Energy-Vector is obtained, vector in this time point will combine the 63 vectors before this time point into a $(64*64)$ feature map. Then the feature extraction step is completed.

B. MODEL CONSTRUCTION

1) CONVOLUTION NEURAL NETWORK

CNN is a feedforward neural network, which is composed of several convolution layers and pooling layers. At present,

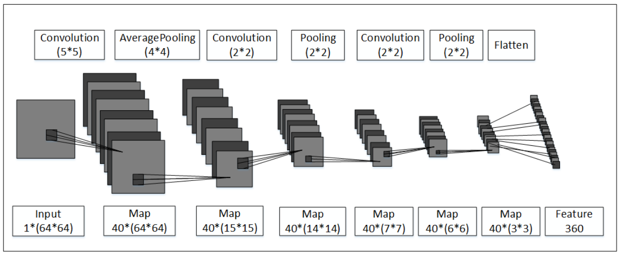


FIGURE 2. **CNN Structure.**

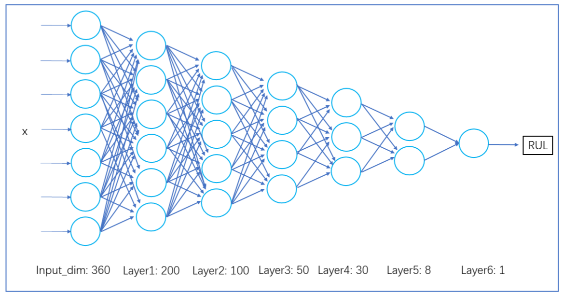


FIGURE 3. **DNN Structure.**

variant CNN has been widely used in image processing. Convolution layer of the CNN has a good perception of the local characteristics of the image, and it can sense the relationship between the pixel and the surrounding pixels [16]. At the same time, the CNN has the characteristics of weight sharing. When the convolution window function is convoluted on the whole feature graph, its parameters remain unchanged, this greatly reduce the number of parameters and reducing the difficulty of training. The pooling layer compresses the convolution results, causing the local features to converge, enabling the discovery of higher levels of law at further convolution [17].

In this paper, the 64 * 64 Spectrum-Principal-Energy-Vector feature map is the input features of the deep CNN. The transverse texture and longitudinal texture are rich, if deep CNN is used to solve the feature map, it can better find the local changes in the law, which can find the relationship between features and the RUL.

The deep convolution neural network (CNN) can extract complex information from the feature map. According to the results of several experiments, this paper chooses the most suitable network structure. The network structure is depicted in figure 2.

The CNN consists of 8 layers, which consists of three convolution layers, three pooling layers, one flatten layer. The network structure is [Convolution Average Pooling Convolution, Dropout, Average Pooling, Convolution, Dropout, Average Pooling]. Since the original feature map contains a lot of detail information, these details are not all necessarily valid for bearing RUL prediction. The function of the network is to find the details from the feature map and to filter useful information from the detail information. In the forward propagation process of the deep CNN, the feature map is gradually blurred, but the overall information of each feature map will be gradually highlighted.The last Flatten layer (fully connected layer) transforms the final information into a feature vector containing 360 elements, and the eigenvector will be used for regression prediction. Each layer uses the ReLU activation function, and the Average Pooling template is 2 * 2.

Rectified linear unit, ReLU, has weakened the drawback of the gradient vanish in training and is currently widely used in convolution networks. The expression is

$$X_{SPEV}(k) = max\{0, x\} \tag{4}$$

2) DEEP NEURAL NETWORK

The CNN model extracts high-level features from the feature map, which can reflect the change of the RUL. But, it still need to be added the regression prediction between the high-level characteristics and the RUL. In many classification and regression problems, the CNN is often followed by several layers of fully connected deep neural networks for classification or regression prediction [9]. In the 6-layer deep neural network, network nodes number of each layer is [200,100,50,30,8,1] and the activation function of each layer is the ReLU function. We followed the previous work to design the network [9].

In this paper, deep neural network (DNN) will be used as prediction model. The structure is depicted in figure 3.

360-Dimensional feature vector obtained by the CNN will be fed into the DNN model , and the RUL is obtained by neural network nonlinear regression prediction.

Another function of the fully connected network is to calculate the prediction error in the training step, the error is the mean square error of the output of the fully connected network and the actual value of the RUL. Through the back propagation of the error, the full connection deep neural network and the deep convolution network layer weight update, to achieve the network learning function.

C. SMOOTHING

The RUL predicted by the prediction network is often not continuous, but the actual bearing RUL is often continuous. This paper uses the forward prediction data to linearly smooth the current forecast data to alleviate the problem of discontinuous predicted RUL.

The relationship between the RUL of the bearing and the running time is linear. The linear regression method is used to smooth the forward prediction result. At the time t, the RUL, Rt, is predicted by the prediction model, and the RUL at ten time-points, $t-9, t-8, t-7, \ldots, t-1, t$, is performed a linear regression. And the regression RUL result at time t will be set as the final predicted value.

The smoothing has almost no impact on prediction error, and has the advantage of the treatment obverse. When long-term monitoring of the bearings, the forecast output is smooth and will not appear fluctuations in the result, and this is in line with the real situation of the bearings RUL.

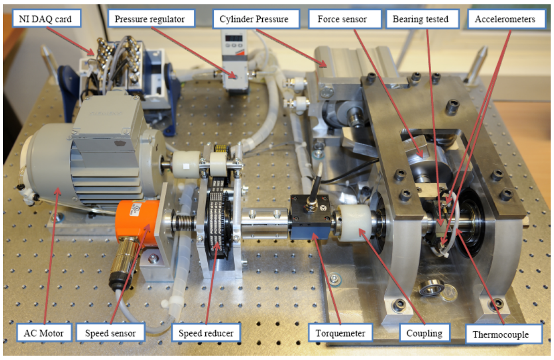


FIGURE 4. PRONOSTIA Experimental Platform [18].

IV. EXPERIMENT AND ANALYSIS

A. DATA DESCRIPTION

The data used in this Experiments are from the IEEE PHM2012 Predictor Challenge experiment data provided by the FEMTO-ST Institute in France [18]. The experimental equipment is shown in Figure 4. The main role of the equipment is to provide experimental data of rolling bearing life. The platform mainly consists of an asynchronous motor, a speed sensor, a temperature sensor, a pressure sensor, and a NIDAQ data acquisition card [18].

The acceleration life test is carried out for each bearing, and the relevant information of the whole life cycle of each bearing is collected, which is mainly the vibration signal in the horizontal direction (X direction) and the vibration signal in the vertical direction (Y direction). All bearing material, specifications, technology, etc. are identical, and there is no initial defect in the bearing. Each bearing failures in the different form, mainly by the material defects and the impact of working conditions. Accelerated life degradation experiments allows the bearing to complete the life cycle degradation process in hours. The time interval for collecting data is 10s, and the time of collecting data is 0.1s [18]. The sampling frequency is 25.6KHz. There will be 2560 sampling points for each sampling. That is, each vibration signal is composed of 2560 Y direction data and 2560 Y direction data, stored in a csv file. The following figure4 shows a bearing full life vibration signal.

B. EXPERIMENT

1) FEATURE MAP CONSTRUCTION

Beginning with vibration signal, The FFT algorithm is used to carry out discrete Fourier transform to obtain the frequency spectrum. And the spectrum is depicted in Figure 5.

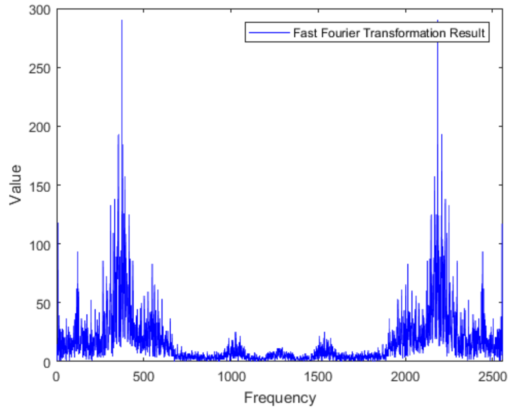


FIGURE 5. Frequency Spectrum.

The original vibration signal has 2560 dimension, and after FFT, the spectrum has 5120 dimension. Due to the symmetry of the spectrum, and the 2560 dimension of the spectrum can reflect all information. The spectrum of the 2560 dimension is divided into 64 blocks, each containing 40 dimensions of spectral information. The maximum amplitude of each spectrum is obtained, and the 64-dimensional Spectrum-Principal-Energy-Vector is obtained. 64 time steps' Spectrum-Principal-Energy-Vectors are combined into a 64*64 feature map. The feature map is depicted in Figure 6.

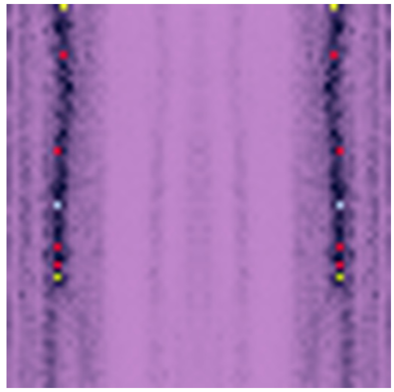


FIGURE 6. Feature Map.

This feature map is the input of the deep CNN. In the figure, the value of black part is small, which means that the energy in the band is low and the white point represents the highest energy. In horizontal view, the change of the texture reflects the difference between the different frequency bands, and in the vertical view, the change of the texture reflects the same frequency band change between different time points. Horizontal and vertical changes all can be used as a basis for the prediction of the RUL. The change of the feature map is complex and rich in texture, which is suitable for the deep CNN.

2) MODEL PREDICTION

The CNN consists of 8 layers. The prediction model is a 6-layer deep neural network, with network parameters of [200,100,50,30,8,1] and the activation function of each layer is the ReLU function. The training loss function for the entire network is 'mean squared error' and the number of epoch is 100.

The test set, never used in training stage, is input into a trained network for RUL prediction, and the results are shown

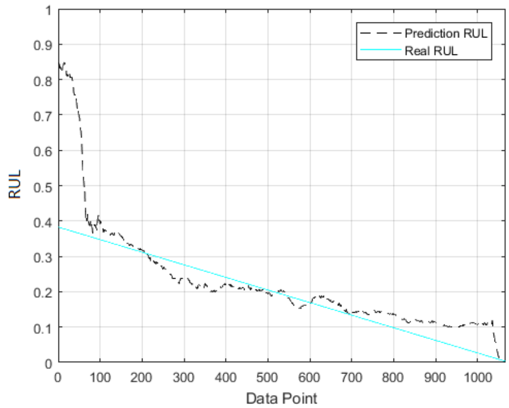


FIGURE 7. Prediction Results.

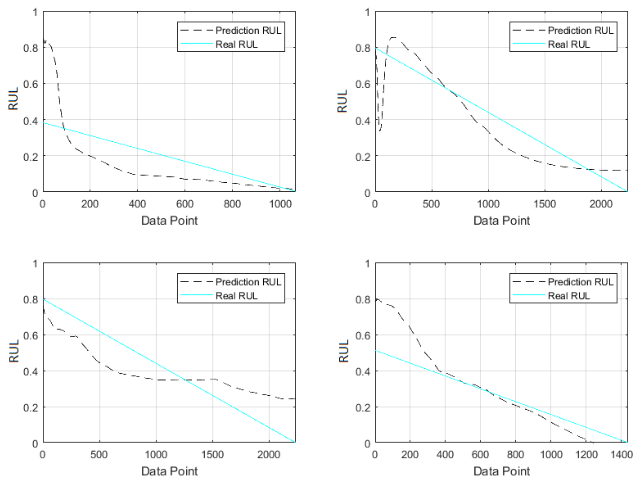


FIGURE 8. Smoothing Results.

in Figure 7. It can be seen from the figure that the prediction results can reflect the trend of bearing degradation.

3) SMOOTHING

It can be seen from the above prediction results that the prediction result is discontinuous.If the prediction results are not smoothed, the RUL fluctuates greatly during the forecasting process. In fact, the RUL should decrease with the use of the bearing. But this phenomenon that the result discontinuous in RUL prediction is common. In order to ensure that the RUL of the bearing is reduced during the whole cycle of the bearing, the prediction results are smoothed with forward data points. Bearings in test set are predicted, and the results are in Figure 8.

The test set contains four bearings with a total mean square error of 0.119. The results show that the method can reflect the degradation trend of bearing performance in the RUL of the bearing, and the method get a good accuracy.

C. COMPARISON AND ANALYSIS

The prediction method of bearing RUL mainly includes two parts: feature extraction method and prediction model. In this paper, a series of comparative experiments are done to verify the validity of the feature and the validity of the model. In the experimental part, the abscissa of each graph is the bearing number, and the ordinate is the mean square error of the predicted result.

1) THE VALIDITY OF THE FEATURE MAP

In this paper, we proposed a feature extraction method for constructing the Spectrum-Principal-Energy-Vector. In order to verify the validity of the feature, the following two groups of experiments were done. The first group of experiments illustrate the reason for each step of the method, and the relevant comparison experiment is done to verify the rationality of the feature extraction method. The second set of experiments compare the Spectrum-Principal-Energy-Vector with the traditional time-frequency domain features to illustrate the effectiveness of the feature.

a: Analysis of feature extraction methods

The extraction of the Spectrum-Principal-Energy-Vector consists of three steps: 1) FFT transform to obtain the signal spectrum information; 2) dividing the spectrum by equidistant segmentation, every 40 discrete frequencies as one frequency band; 3) taking the maximum value for each frequency band.

In the above process, the first step to transform the signal to the frequency domain. The vibration signal is chaotic, and the noise information is much [13]. Compression process will lose the frequency information. And in the frequency domain, it will be ensured that the low-frequency signal and noise will be separated. In addition, the spectral amplitude information can reflect the original vibration signal amplitude.

In the third step, the maximum value is taken for each frequency band, because the maximum energy of the point has the highest amplitude in the spectrum. The spectrum of the middle frequency and low frequency band contains low noise. So taking the maximum amplitude will not be affected by noise too much. On the contrary, in the original vibration signal, taking the maximum value will be a huge noise interference. In addition, between the maximum values and mean values of the spectrum, it can be clearly seen that the mean information makes the variation between the features too smooth and cannot reflect the change information well.

In order to verify the Spectrum-Principal-Energy-Vector suitable for the proposed model, this paper changes the first and third steps of the feature extraction method compared with Spectrum-Principal-Energy-Vector. We use two kinds of features to compare with the Spectrum-Principal-Energy-Vector, and use the Amplitude Mean Energy Vector extracted from the original signal and the Spectral Mean Energy Vector. These two feature extraction methods are similar to the Spectrum-Principal-Energy-Vector. The difference is that the Spectral Mean Energy Vector uses the mean method when compressing the 40-dimensional spectrum. The Amplitude Mean Energy Vector is extracted from the original vibration signal and is averaged using a 40-bit signal for compression.

The results are shown in the Figure 9. The following figure shows the mean square error of the four bearings

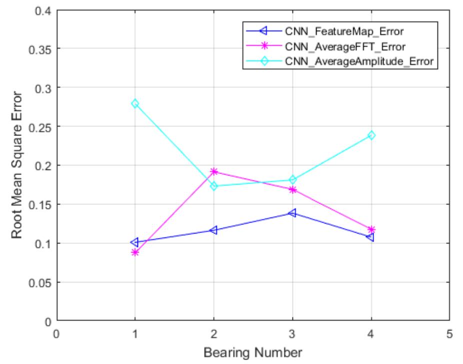


FIGURE 9. Extraction Methods Comparison.

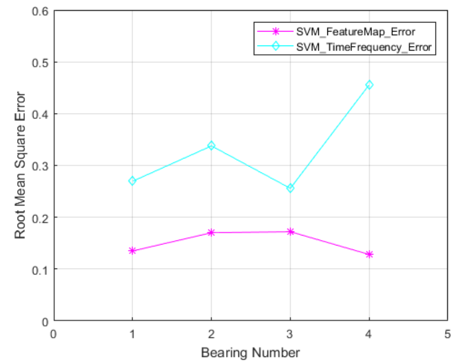


FIGURE 10. Feature Comparison in SVM.

TABLE 1. Extraction method result.

Feature	RMSE
Spectrum-Principal-Energy-Vector(Blue line)	0.119
Spectral Mean Energy Vector(Red line)	0.157
Amplitude Mean Energy Vector(Green line)	0.209

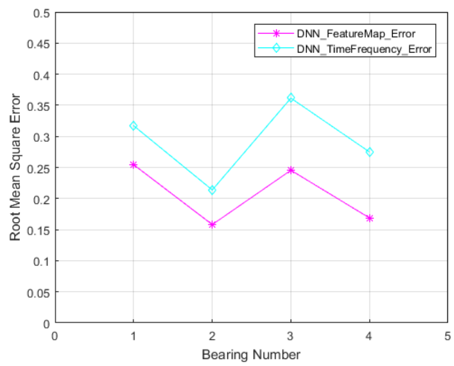


FIGURE 11. Feature Comparison in DNN.

predicted lives. The four bearings are the horizontal axis [7], [13], [19]. CNN_FeatureMap_Error (Blue line) represents the Spectrum-Principal-Energy-Vector prediction result used in this paper. CNN_AverageFFT_Error (Red line) represents the result of Spectral Mean Energy Vector. CNN_AverageAmoplitude_Error (Green line) represents the result of Amplitude Mean Energy Vector.

The total error of the four sets of bearings is shown in the following table 1. In the table, RMSE means Root Mean Squared Error. It can be seen from the above table and the graph that the prediction error of the Spectrum-Principal-Energy-Vector(Blue line)is the lowest, which proves that the Spectrum-Principal-Energy-Vector(Blue line) is suitable for the RUL prediction of bearing in deep CNN. In addition, the prediction error generated by the AverageFFT feature is lower than the AverageAmplitude feature as a whole, and it can be seen that the frequency domain feature has a greater advantage over the time domain features in the bearing useful life prediction.

b: Contrast with traditional features

In order to verify the validity of the feature, this paper compares the classical time-frequency domain features [13] with the Spectrum-Principal-Energy-Vector.

The Support Vector Machines (SVM) [20] and the Deep Neural Network (DNN) are used as the prediction model to test the validity of the two features respectively. The two models are used to illustrate that the Spectrum-Principal-Energy-Vector has a strong applicability. Figure 10 uses the SVM model to compare the time-frequency characteristics with the Spectrum-Principal-Energy-Vector.

Figure 11 uses the DNN model to compare the time-frequency characteristics with the Spectrum-Principal-Energy-Vector.

It can be seen from the results of the above two figures that the prediction effect of the Spectrum-Principal-Energy-Vector is better than that of the time-frequency domain features when using the SVM and the deep neural network as the prediction model. It can be deduced that the Spectrum-Principal-Energy-Vector feature has a wide range of applicability. At the same time, when using the Spectrum-Principal-Energy-Vector feature map to predict, SVM model error between the different bearings change little, and DNN model error changes seriously. It can be seen that DNN model is over-fitting with high dimension input.

2) THE VALIDITY OF THE CNN MODEL

In order to verify the effectiveness of the CNN model, the following two groups of experiments were done. The first group of experiments uses the same feature input, compared the effect of different models to verify the effectiveness of the CNN model. However, since the same feature is not necessarily suitable for different models, the second experiments uses the feature that are compatible with each model to verify the effectiveness of the CNN model.

a: Model Comparison, with same feature

In this experiment, the deep neural network model and SVM model are compared with the CNN model. The input of three models is consistent, the Spectrum-Principal-Energy-Vector. The results are shown in Figure 12.

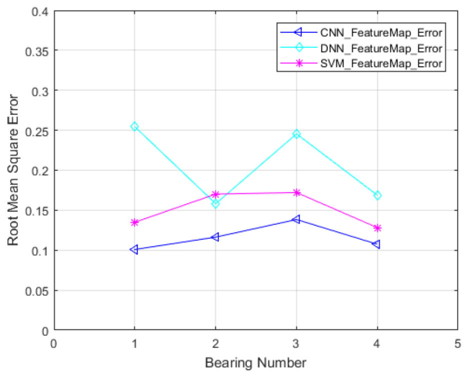


FIGURE 12. Model Comparison with Same Feature.

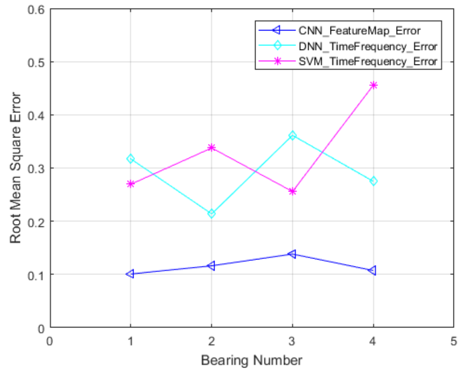


FIGURE 13. Model Comparison with different Feature.

The blue line in the figure shows the results of the deep CNN model. In the three models, the deep fully connected neural network is the worst, mainly due to the large input scale and the serious over-fitting problem. Due to the existence of the mechanism of weight sharing, the same input size and network layer, the deep CNN parameters are less than the deep fully connected neural network [8], [21], largely avoiding the over-fitting problem. At the same time, the support vector machine model is much lower than the deep fully connected neural network, which is more suitable for small-scale data learning, and has some ability to grasp the law of the RUL. However, the SVM model lacks the ability to find deep-seated laws. So, the SVM model is worse than the deep CNN.

It can be seen that the predictive effect of each bearing by the deep CNN is better than that of the other two models, and the validity of the deep CNN model is verified.

3) MODEL COMPARISON, WITH DIFFERENT FEATURES

In the experiment with same feature, there may be cases where the model does not match the features. Some classical models may only be suitable for classical features. So, in this experiment, the input of the classical model is classical feature, and the result will be compared with the method proposed in this paper. In this experiment, time-frequency features by wavelet transform will be set as the input of SVM and DNN model. The results are shown in Figure 13 and Table 2:

TABLE 2. Model result.

Feature	RMSE
CNN with Spectrum-Principal-Energy-Vector(Blue line)	0.119
SVM with time-frequency features(Red line)	0.334
DNN with time-frequency features(Green line)	0.296

It can be seen from the above experiments that the deep CNN model used in this paper can improve the prediction accuracy of bearing RUL.

From the above series of experiments, the feature extraction method of the Spectrum-Principal-Energy-Vector can improve the accuracy significantly for the bearing RUL prediction. The prediction error is reduced from 0.334 to 0.16 in SVM. In addition, the prediction model based on deep CNN can further improve the prediction accuracy, and the prediction accuracy can reach 0.1190. It can be seen that the Spectrum-Principal-Energy-Vector proposed in this paper can have a better representation ability on the original data. At the same time, the prediction model based on deep CNN proposed in this paper can improve the prediction accuracy.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a new method based on deep convolution neural network for the prediction of bearing RUL. Our analysis and experiments using real-world data have shown significantly improved prediction accuracy. Our new feature extraction method, the Spectrum-Principal-Energy-Vector can better represent the information from the raw data. This eigenvector can be combined with different prediction models in different scenarios for different types of data in both time and frequency domains. For the construction of prediction models, we propose a new scheme based on deep convolution neural network. In the stage of RUL prediction, we present a post-smoothing method to address the discontinuity problem in the prediction results, greatly improve the interpretability of the prediction results.

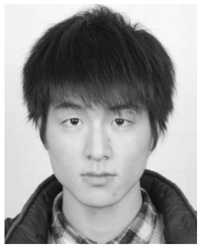
We plan to further improve the feature extraction by refining the spectrum segmentation from 64 to 1 for our Spectrum-Principal-Energy-Vector. We will also explore other deep learning algorithms such as *Recurrent Neural Networks, RNN,* for its advantages in time sequential data processing, and will combine CNN with RNN to improve the accuracy.

REFERENCES

[1] X. Wang, L. T. Yang, X. Xie, J. Jin, and M. J. Deen, "A cloud-edge computing framework for cyber-physical-social services," *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 80–85, Nov. 2017.
[2] R. Razavi-Far, M. Farajzadeh-Zanjani, and M. Saif, "An integrated class-imbalanced learning scheme for diagnosing bearing defects in induction motors," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 2758–2769, Dec. 2017.
[3] K. Zhu and T. Liu, "Online tool wear monitoring via hidden semi-Markov model with dependent durations," *IEEE Trans. Ind. Informat.*, vol. 14, no. 1, pp. 69–78, Jan. 2018.
[4] Y. Wang, Y. Peng, Y. Zi, X. Jin, and K. L. Tsui, "A two-stage data-driven-based prognostic approach for bearing degradation problem," *IEEE Trans. Ind. Informat.*, vol. 12, no. 3, pp. 924–932, Jun. 2016.
[5] M. Grasso, S. Chatterton, P. Pennacchi, and B. M. Colosimo, "A data-driven method to enhance vibration signal decomposition for rolling bearing fault analysis," *Mech. Syst. Signal Process.*, vol. 81, pp. 126–147, Dec. 2016.

[6] W. Li, S. Zhang, and S. Rakheja, "Feature denoising and nearest–farthest distance preserving projection for machine fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 12, no. 1, pp. 393–404, Jan. 2016.
[7] T. H. Loutas, D. Roulias, and G. Georgoulas, "Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic E-support vectors regression," *IEEE Trans. Rel.*, vol. 62, no. 4, pp. 821–832, Dec. 2013.
[8] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[9] L. Ren, J. Cui, Y. Sun, and X. Cheng, "Multi-bearing remaining useful life collaborative prediction: A deep learning approach," *J. Manuf. Syst.*, vol. 43, pp. 248–256, Apr. 2017.
[10] J. Hu, L. Shen, and G. Sun. (Sep. 2017). "Squeeze-and-excitation networks." [Online]. Available: https://arxiv.org/abs/1709.01507
[11] R. K. Singleton, E. G. Strangas, and S. Aviyente, "Time-frequency complexity based remaining useful life (RUL) estimation for bearing faults," in *Proc. IEEE Int. Symp. Diagnostics Electr. Mach., Power Electron. Drives*, Aug. 2013, pp. 600–606.
[12] S. Dong and T. Luo, "Bearing degradation process prediction based on the PCA and optimized LS-SVM model," *Measurement*, vol. 46, no. 9, pp. 3143–3152, 2013.
[13] K. Medjaher, D. A. Tobon-Mejia, and N. Zerhouni, "Remaining useful life estimation of critical components with application to bearings," *IEEE Trans. Rel.*, vol. 61, no. 2, pp. 292–302, Jun. 2012.
[14] A. Ragab, M.-S. Ouali, S. Yacout, and H. Osman, "Remaining useful life prediction using prognostic methodology based on logical analysis of data and Kaplan–Meier estimation," *J. Intell. Manuf.*, vol. 27, no. 5, pp. 943–958, 2016.
[15] E. V. Zaretsky, "Rolling bearing life prediction, theory, and application," NASA Glenn Res. Center, Cleveland, OH, USA, Tech. Rep. NASA/TP-2013-215305 E-16300-1, 2013.
[16] K. Simonyan and A. Zisserman. (Sep. 2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[18] P. Nectoux *et al.*, "PRONOSTIA: An experimental platform for bearings accelerated life test," in *Proc. IEEE Int. Conf. Prognostia Health Manage.*, Denver, CO, USA, Jun. 2012, pp. 1–8.
[19] A. Soualhi, H. Razik, G. Clerc, and D. D. Doan, "Prognosis of bearing failures using hidden Markov models and the adaptive neuro-fuzzy inference system," *IEEE Trans. Ind. Electron.*, vol. 61, no. 6, pp. 2864–2874, Jun. 2013.
[20] N. Gebraeel, M. Lawley, R. Liu, and V. Parmeshwaran, "Residual life predictions from vibration-based degradation signals: A neural network approach," *IEEE Trans. Ind. Electron.*, vol. 51, no. 3, pp. 694–700, Jun. 2004.
[21] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

LEI REN (M'17) received the Ph.D. degree in computer science from the Institute of Software, Chinese Academy of Sciences in 2009. He is currently an Associate Professor and the Deputy Head with the Cloud Manufacturing Research Center, School of Automation Science and Electrical Engineering, Beihang University, China. He has authored or co-authored over 60 papers and got over 2000 citations according to Google Scholar. He edited a book *Challenges and Opportunity with Big Data* published by Springer LNCS. His research interests include cloud manufacturing, industrial Internet of Things, and big data analytics. He served as an Associate Editor of *SIMULATION: Transactions of the Society for Modeling and Simulation International*, and reviewers for journals, such as IEEE TII, TSMC, and ACM TCC. He also served as a TPC Co-Chair for the 19th Monterey Workshop on Big Data and Fifth International Conference on Enterprise Systems. He is a member of the IEEE Computer Society and IEEE Systems, Man, and Cybernetics Society.



YAQIANG SUN received the B.Eng. degree in information engineering from Beihang University, China, in 2016, where he is currently pursuing the Ph.D. degree with the School of Automation Science and Electrical Engineering. His research interests include deep learning, industrial big data analytics, and cloud-based prognosis for manufacturing.

HAO WANG (M'12) He received the B.Eng. and Ph.D. degrees in 2000 and 2006, respectively, both in computer science. He was a Researcher with IBM Canada, McMaster, and St. Francis Xavier University. He is currently an Associate Professor and the head with the Big Data Laboratory, Department of ICT and Natural Sciences, Norwegian University of Science and Technology, Norway. He has authored or co-authored over 60 papers in IEEE TVT, GLOBALCOM 2016, Sensors, IEEE Design and Test, and Computer Communications. His research interests include big data analytics and industrial Internet of Things, high performance computing, safety-critical systems, and communication security. He served as a TPC Co-Chair for the IEEE DataCom 2015, the IEEE CIT 2017, and ES 2017, and reviewers for journals, such as IEEE TKDE, TBD, TETC, T-IFS, and ACM TOMM. He is a member of IEEE IES Technical Committee on Industrial Informatics.

LIN ZHANG (M'03–SM'11) received the B.S. degree from the Department of Computer and System Science, Nankai University, China, in 1986, and the M.S. and Ph.D. degrees from the Department of Automation, Tsinghua University, China, in 1989 and 1992, respectively. From 2002 to 2005, he was with the U.S. Naval Postgraduate School, U.S. National Research Council, as a Senior Research Associate. He is currently a Full Professor with Beihang Unversity. He has authored or co-authored 200 papers, ten books and chapters. His research interests include service-oriented manufacturing and simulation, cloud manufacturing, and model engineering for simulation. He was a recipient of the National Award for Excellent Science and Technology Books in 1999, the 863 Outstanding Individual Award in 2001, and the National Excellent Scientific and Technological Workers Awards in 2014. He serves as the Past President of the Society for Modeling and Simulation International, a Vice President of the Chinese Simulation Federation, a Fellow of ASI-ASIM, a Board Member of CAAI, a member of the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, the Chief Scientist of key projects of China High-Tech Research and Development Program (863), and the Associate Editor-in-Chief and an Associate Editor of six peer-reviewed international journals.