Bearing degradation trend prediction under different operational conditions based on CNN-LSTM

Guozeng Liu*, Jianmin Zhao and Xin Zhang
Army Engineering University, Shijiazhuang, China.

*Corresponding author e-mail: xiaoliu2005@sohu.com

Abstract. Bearing degradation research is an important part in condition-based maintenance. This paper tries to propose an end-to-end model to realize bearing degradation trend prediction by vibration signals. Convolution neural network is good at data dimension reduction and feature extraction, and long short-term memory is good at deal with time sequences. The experiment proves that the combination of the two deep learning methods has a good effect, which avoid some disadvantages of traditional methods. Results shows that the model has application value in industrial practice.

1. Introduction

Bearings are one of the most important parts in industrial production. The failure of bearings will lead to the loss of economy or even the casualties. Thus, the research of bearing health attracts much attention. And bearing degradation trend prediction is a part of condition-based maintenance[2].

The study of bearing degradation is mainly based on vibration signals. Traditional methods usually extract some features from vibration signals as degradation index and then establish models to realize degradation trend prediction. Time domain analysis, frequency analysis and time-frequency domain analysis are the most widely used feature extraction methods, including Fourier transform[6], wavelet transform (WT)[8], empirical mode composition (EMD)[12] and some other improved methods. Those are able to make a detailed analysis of the signal. After getting the features, some models are applied to perform bearing degradation. Statistical models are usually used in this field, such as hidden Markov model (HMM)[5], Bayes model[1], Gaussian mixture model (GMM)[7] and so on. The other kind of model is machine learning models like support vector regression (SVR)[3], neural networks (NN)[9] etc.

The combination of traditional methods above has developed a lot. However, they have some disadvantages hard to avoid. One is that it is difficult for them to predict the degradation trend when the operational conditions change. Most methods have to pretreat signals to eliminate the influence of speed and load variety and it will lead to the loss of information[4,13]. The other is that the combination of methods will lead to the accumulation of error in the steps. Deep learning is a rising technique which can learn the features of data adaptively and establish an end-to-end model. Here are some common methods of deep learning like deep belief network (DBN)[11], convolutional neural network (CNN)[10], recurrent neural network (RNN), and some other neural network methods.

This paper tries to make a model which combines CNN and long short-term memory (LSTM). CNN is able to provide degradation features of signals and LSTM can memorize the time series information.
Experiments shows that the model is superior in accuracy and efficiency when dealing with the degradation trend prediction of different operational conditions.

2. Technical background

2.1. Convolutional neural network

CNN is a typical neural network of deep learning. It is inspired by biological processes in which the connectivity pattern among neurons resembles the organization of the animal visual cortex. So, it is good at dealing with the large amount of data. The crucial function of CNN is learning the features of signals and reduce the redundant information in every sample. CNN is made up of input layer, convolution layers, pooling layers, full connection layers and output layer.

In CNN, convolution layers and pooling layers are the most important parts which take charge of degradation feature learning. The parameters of convolutional layers contain kernel size, kernel numbers, stride etc. By convolving the upper input matrix with different convolution kernels, the local features are combined to form a convolution kernel feature matrix. The convolution operation calculation expression is as shown in equation (1):

$$x_{l}^{j} = f\left( \sum_{i \in M_{l}} x_{l-1}^{j} \times k_{l}^{j} + b_{l}^{j} \right)$$

where $M_{l}$ is the input vector, $l$ is the $l$th layer, $k$ is convolutional kernel, $b$ is bias vector, $x_{l}^{j}$ is the output of the $l$th layer and $x_{l-1}^{j}$ is the input of the $l$th layer. The pooling layer is a down sampling on the feature vector of convolutional layers. It uses the local correlation principle of the feature to perform aggregation statistics in adjacent small regions to extract more important feature information further. Moreover, different lengths of signals input can generate fixed-dimensional feature vectors through the pooling layers and pass the pooled layer output to the fully-connected layer for classification. Commonly used down sampling operations include average pooling, max pooling, and stochastic pooling. Among them, max pooling is the most widely used. The calculation expression of the pooling operation is as shown in equation (2):

$$P_{l}^{(j)} = \max_{(j-1)r+c \in (j)r} \{ q_{l}^{(m)} \}$$

In order to enhance the nonlinear mapping capability of the network and limit the size of the network, the network accesses a fully connected layer after extracting features from the feature extraction layers. In regression problem, the model just has single input and single output. Each neuron in this layer is interconnected with all neurons in the previous layer, and the neurons in the same layer are not connected. The example of CNN structure is shown in Figure 1.

![Figure 1. An example of CNN](image-url)
2.2. Long short-term memory
LSTM is a kind of RNN, which is a type of neural network for processing sequence data. RNN contains loops, allowing information to be persisted. It can be used to connect previous information to the current task. However, when the interval between the related information and the current predicted position becomes larger, RNN will lose the ability to learn to connect to such far away information. LSTM is proposed to solve the problem. A common architecture is composed of a cell and three "regulators", usually called gates, of the flow of information inside the LSTM unit: an input gate, an output gate and a forget gate. Some variations of the LSTM unit do not have one or more of these gates or maybe have other gates. The compact forms of the equations for the forward pass of an LSTM unit with a forget gate are shown in equation (4):

\[
\begin{align*}
    f_t &= \sigma_x(W_f x_t + U_f h_{t-1} + b_f) \\
    i_t &= \sigma_x(W_i x_t + U_i h_{t-1} + b_i) \\
    o_t &= \sigma_x(W_o x_t + U_o h_{t-1} + b_o) \\
    c_t &= f_t c_{t-1} + \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \\
    h_t &= o_t \sigma_h(c_t)
\end{align*}
\]

where \( x_t \) is input vector to the LSTM unit, \( f_t \) is forget gate's activation vector, \( i_t \) is input gate's activation vector, \( o_t \) is output gate's activation vector, \( h_t \) is hidden state vector also known as output vector of the LSTM unit, \( c_t \) is cell state vector, \( \sigma_x \) is sigmoid function, \( \sigma_c \) is hyperbolic tangent function, \( \sigma_h \) is hyperbolic tangent function or linear function, \( W, U \) and \( b \) are weight matrices and bias vector parameters and the operator \( \circ \) denotes the Hadamard product.

Figure 2 gives the structure of LSTM. The sigmoid layer is called the "input gate layer" and determines what value we are going to update. Then, the tanh layer creates a new candidate value vector, \( c_t \), that will be added to the state. In other words, the sigmoid function selects the update content, and the tanh function creates the update candidate.

3. Experiment

3.1. Model construction
Bearing degradation trend prediction is essentially a time series prediction problem. In the end-to-end model, when the whole vibration signal data is used as input, the difficulty of LSTM training will increase a lot. CNN is introduced to avoid the trouble, which can reduce the original data dimension and provide the features. At first, the run-to-failure data is sampled once in a while and it will be divided into some subsequences. Then, every three time-steps are set as input to the convolution layer and pooling layer. The features from CNN will be packaged into LSTM. The model is shown in figure 3.
In the CNN part, the parameters of the model include filters, kernel size, stride and pool size, which influence the correlation of features and dimension reduction efficiency. In LSTM part, the number of neurons in hidden layer, units, is the crucial parameter, which determines the ability to express degradation trends. Some pre-tests are carried on to find the ideal parameters. Table 1 gives the value of the parameters.

Table 1. The value of the parameters.

| Parameter | Filters | Kernel size | Strides | Pool size | Units |
|-----------|---------|-------------|---------|-----------|-------|
| Value     | 6       | 110         | 10      | 19        | 20    |

3.2. Data analysis

The bearing degradation vibration data under different operational conditions comes from FEMTO-ST Institute in PHM challenge 2012. The vibration data is collected at a sampling frequency of 25.6 KHz every 10 s for 0.1 s. The bearing is considered to be invalid, when the acceleration amplitude exceeds 20g. Here are 3 kind of operational conditions in total and each operational condition contain 2 bearings. The training samples contain the vibration data from every operational condition and mix different operational conditions. The rest composes test samples. The details are shown in table 2.

Table 2. The operational conditions of bearings.

| Sample kind | Bearing    | Load(N) | Speed(rpm) | Time(min) |
|-------------|------------|---------|------------|-----------|
| Training samples | Bearing1-1 | 4000    | 1800       | 467       |
|              | Bearing2-1 | 4200    | 1650       | 151.67    |
|              | Bearing3-1 | 5000    | 1500       | 65.67     |
|              | Bearing1-2 | 4000    | 1800       | 145       |
| Test samples | Bearing2-2 | 4200    | 1650       | 132.67    |
|              | Bearing3-2 | 5000    | 1500       | 272.67    |

Taking root mean square (RMS) as a prediction target, the comparison of real and predicted values are shown in figure 4.
To evaluate the effect of the model in this paper, mean squared error (MSE) and mean absolute error (MAE) are used to compare with other methods like CNN, WT-SVR and EMD-GMM. The comparison is shown in table 3 taking bearing 1-2 as an example.

Table 3. Three Scheme comparing.

| Scheme                  | MSE      | MAE      | Training time (s) |
|-------------------------|----------|----------|-------------------|
| CNN-LSTM                | 0.0010   | 0.0203   | 174.13            |
| CNN                     | 0.0217   | 0.2352   | 136.21            |
| WT-SVR                  | 0.0103   | 0.1949   | 374.82            |
| EMD-GMM                 | 0.0530   | 0.4668   | 706.90            |

Conclusion

Bearing degradation problem is more complex when the operational conditions are different. This paper combines CNN with LSTM to finish an end-to-end model, which is able to predict bearing degradation trend. It can avoid the accumulation of error in multi-step method and the bad influence among different steps. From the experiment result, CNN-LSTM is superior to traditional methods in efficiency and accuracy. It has high value in industrial application.

References

[1] Addin, O., Sapuan, S. M., Mahdi, E., and Othman, M., A Naïve-Bayes classifier for damage detection in engineering materials, Materials & Design 28 (2007) 2379–2386.
[2] Al-Najjar, B., Condition-based maintenance: Selection and improvement of a cost-effective vibration-based policy in rolling element bearings, PhD Thesis, KFS AB, Lund, (1997).
[3] Benkedjouh, T., Medjaher, K., Zerhouni, N., and Rechak, S., Remaining useful life estimation based on nonlinear feature reduction and support vector regression, Engineering Applications of Artificial Intelligence 26 (2013) 1751–1760.
[4] Borghesani, P., Pennacchi, P., Randall, R. B., Sawalhi, N., and Ricci, R., Application of cepstrum pre-whitening for the diagnosis of bearing faults under variable speed conditions, Mechanical Systems and Signal Processing 36 (2013) 370–384.
[5] Boutros, T., and Liang, M., Detection and diagnosis of bearing and cutting tool faults using hidden Markov models, Mechanical Systems and Signal Processing 25 (2011) 2102–2124.
[6] Bracewell, R. N., and Bracewell, R. N., The Fourier transform and its applications, (McGraw-Hill New York, 1986).
[7] Nelwamondo, F. V., Marwala, T., and Mahola, U., Early classifications of bearing faults using hidden Markov models, Gaussian mixture models, mel-frequency cepstral coefficients and fractals, International Journal of Innovative Computing, Information and Control 2 (2006) 1281–1299.
[8] Pan, Y., Chen, J., and Guo, L., Robust bearing performance degradation assessment method based on improved wavelet packet-support vector data description, Mechanical Systems and Signal Processing 23 (2009) 669–681.
[9] Ren, L., Cui, J., Sun, Y., and Cheng, X., Multi-bearing remaining useful life collaborative prediction: A deep learning approach, Journal of Manufacturing Systems 43 (2017) 248–256.

[10] Sateesh Babu, G., Zhao, P., and Li, X.-L., Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life, in Database Systems for Advanced Applications, eds. Navathe, S. B., Wu, W., Shekhar, S., Du, X., Wang, X. S., and Xiong, H., (Springer International Publishing, 2016), pp. 214–228.

[11] Shao, H., Jiang, H., Zhang, X., and Niu, M., Rolling bearing fault diagnosis using an optimization deep belief network, Meas. Sci. Technol. 26 (2015) 115002.

[12] Soualhi, A., Medjayer, K., and Zerhouni, N., Bearing Health Monitoring Based on Hilbert–Huang Transform, Support Vector Machine, and Regression, Instrumentation and Measurement, IEEE Transactions On 64 (2015) 52–62.

[13] Teotakool, K., Devaney, M. J., and Eren, L., Adjustable-Speed Drive Bearing-Fault Detection Via Wavelet Packet Decomposition, IEEE Transactions on Instrumentation and Measurement 58 (2009) 2747–2754.