Rolling bearing diagnosis method based on WMKPCA-LSTM network

Bo Cui, Jizheng Chu* and Ning Cao
College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, 100010, China
Jizheng Chu E-mail: chujz@mail.buct.edu.cn

Abstract. The monitoring data from large machinery have characteristics of high dimensions, non-linearity and correlation. In the traditional fault diagnosis, it is hard to extract features and complex to build the model. In order to solve these problems, the rolling bearing diagnosis method based on a weighted mixed kernel principal component analysis (WMKPCA) and long-short-term memory (LSTM) network is proposed. The method uses WMKPCA algorithm to reduce the dimensionality of the vibration signal. The method not only effectively reduces the correlation of the original data, but also improve the diagnostic accuracy; The LSTM neural network has a memory function that can be used to extract temporal features. Experimental results show that the proposed algorithm has higher diagnostic accuracy and higher efficiency than traditional methods of principal component analysis (PCA), kernel principal component analysis (KPCA) and multiway kernel principal component analysis (MKPCA).

1. Introduction

Industrial production technology has greatly improved the development of productivity. As one of the key equipment of industrial production, rotary machinery has the characteristics of huge volume, complicated structure and various parts. Once the rotary machinery fails, it will bring huge economic losses to enterprises and countries. As a key part of large rotating machinery, rolling bearing is one of the most vulnerable industrial parts. According to statistics, rolling bearing failures account for about 30% of all rotating machinery failures [1]. Therefore, it is particularly important to realize fault diagnosis and status recognition of rolling bearings for large enterprises. In recent years, with the increase of the number of mechanical devices, measuring points and sampling frequency, fault diagnosis of mechanical equipment has entered the era of "big data" [2]. The vibration signal generated by the rolling bearing has the characteristics of nonlinearity and redundancy. How to use "big data" to improve the accuracy of fault diagnosis will become a research hotspot in the future.

The vibration signal, as an important reliability indicator, can demonstrate the dynamic evolution of faults and has been widely used in data-driven fault diagnosis methods. According to the characteristics of vibration signals, time domain analysis, frequency domain analysis, and time-frequency analysis are proposed. However, these methods have many defects, such as numerous parameters, high dimensions and redundant information. The PCA was an effective method for analyzing data [3]. It could map high-dimensional data to low-dimensional feature space and retain the feature quantity with large information contribution, but wasn’t suitable for nonlinear data [4]. The literature [5] improved classification accuracy with the KPCA algorithm when the number of principal elements was less than linear PCA. In [6], a new fault diagnosis method based on KPCA-PSVM is proposed. This method can effectively capture the nonlinear relationship between variables, and the accuracy is obviously better than PCA-PSVM. In
[7], a multi-channel KPCA fault detection method was proposed, which improved the feature quality and was suitable for processing complex data. The traditional MKPCA algorithm only combines the weights of local kernel functions and global kernel functions, but the focus of the two kernel functions for extracting data features is different. The above-mentioned fault diagnosis methods have obtained good results on small sample data. However, as the amount of monitoring data increases, these methods no longer applies. Deep learning is a special kind of machine learning with powerful feature extraction capabilities [8-10]. Therefore, it is widely applied to bearing fault diagnosis.

This paper proposes a fault diagnosis method of rolling bearings based on the WMKPCA algorithm and LSTM network. The method uses the WMKPCA algorithm to reduce the dimensionality of the data and extract the data features, and uses the LSTM network to implement feature learning and fault classification. The WMKPCA algorithm can eliminate redundant information in the data and adaptively adjust the local kernel and global kernel weight. Therefore, the WMKPCA algorithm can take full advantage of the two kernel functions. The LSTM network has a memory function and can extract the time dimension features of the data [11].

2. The weighted mixed kernel principal component analysis

The PCA has the function of dimension reduction and feature extraction in fault diagnosis. However, PCA is a linear method that does not perform well in nonlinear processes. Various nonlinear methods have been proposed for detecting the nonlinear characteristics of data, such as KPCA. The KPCA method uses the kernel function to map the original feature space of the sample to the high-dimensional feature space, so that the sample can be linearly separable at the high-dimensional space, and then uses the PCA algorithm to reduce the dimension. Compared with PCA, KPCA can more effectively retain the characteristics of the original data and maximize the nonlinear information in the data.

In the KPCA algorithm, there are mainly four kinds of kernel functions, namely, linear kernel function, polynomial kernel function, radial basis function kernel, and sigmoid kernel function. These kernel functions are divided into two types: global kernel function and local kernel function. The former has weak learning ability and strong generalization ability, while the latter has strong learning ability and weak generalization ability. This paper briefly analyzes the radial basis function (RBF) kernel and the polynomial (Poly) kernel function.

- Local kernel function

As a typical local kernel function, the RBF kernel function has a strong ability to extract local features of samples. The expression is as follows:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$ (1)

- Global kernel function

A typical representation of the global kernel function is the Poly kernel function. The expression is as follows:

$$K(x_i, x_j) = \left(\tilde{x}_i \cdot \tilde{x}_j + 1\right)^d, \quad d \in \mathbb{Z}^+$$ (2)

The performance of the single kernel function method is not ideal. The Poly kernel function has strong generalization ability and the RBF kernel function has strong learning ability. Therefore, the WKPCA is further proposed. The expression is as follows:

$$K'(x_i, x_j) = \lambda K_{RBF}(x_i, x_j) + (1 - \lambda)K_{Poly}(x_i, x_j)$$

$$= \lambda \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) + (1 - \lambda)\left(\tilde{x}_i \cdot \tilde{x}_j + 1\right)^d$$ (3)

Its mixed kernel matrix formula is as follows:
\[
K(x_i, x_j) = \begin{cases}
(1 - \lambda \left(\|x_i\|^2 + 1\right))^d, & x_i = x_j \\
K(x_i, x_j), & x_i \neq x_j
\end{cases}
\]  

(4)

Where \( x_i \) and \( x_j \) are n-dimensional input data; \( K_{\text{RBF}}, K_{\text{Poly}} \) represents the RBF kernel function and the Poly kernel function, respectively; \( \sigma \) is the parameter of the RBF kernel function; \( d \) is the order of the Poly kernel function; \( \lambda \in [0, 1] \) represents the balance parameter. The balance parameter \( \lambda \) is used as a hyperparameter in the model training so that the global and local features of the mixed kernel can be dynamically balanced.

3. Long short-term memory neural network

Recurrent Neural Networks (RNN) is an artificial neural network. Compared with traditional neural networks, RNN has a memory function and is suitable for processing time series data. However, the RNN has the problems of gradient disappearance, gradient explosion and loss of historical information. The LSTM neural network is an improved network of the RNN and solves the above problems of RNN by adding three gate units and memory unit. The state transition equation is follows: \( h_{t-1}^i, c_{t-1}^i \rightarrow h_t^i, c_t^i \) and its structure is as shown:

![LSTM neural network](image)

The output of the LSTM network is determined by the output gate, the input gate, the forgetting gate, and the memory unit values at the previous moment. The specific implementation process is as follows:

- The expression of forget gate \( f_t \), input door \( i_t \) and output door \( o_t \) are as follows:
  \[
  f_t = \sigma \left( W_f \cdot [h_{t-1}, x_t] + b_f \right) 
  \]  
  \[ (5) \]
  \[
  i_t = \sigma \left( W_i \cdot [h_{t-1}, x_t] + b_i \right) 
  \]  
  \[ (6) \]
  \[
  o_t = \sigma \left( W_o \cdot [h_{t-1}, x_t] + b_o \right) 
  \]  
  \[ (7) \]

The forget gate saves useful information of the previous memory unit value; the input gate is used to save useful information of the current input value; the output gate determines the current output value.

- The current memory state unit is determined by the memory state unit candidate value, the input gate and the forgetting gate. The formula is as follows:
  \[
  \tilde{C}_t = \tanh \left( W_c \cdot [h_{t-1}, x_t] + b_c \right) 
  \]  
  \[ (8) \]
  \[
  C_t = f_t \cdot \tilde{C}_{t-1} + i_t \otimes \tilde{C}_t 
  \]  
  \[ (9) \]

Finally, the current output is calculated according to the output gate and the current memory unit value.

\[
  h_t = o_t \cdot \tanh \left( C_t \right) 
  \]  

(10)

Where \( x_t \) and \( h_t \) are input and output data at time \( t \). \( i_t, f_t, o_t \) are input gates, forget gate and output gate, \( C_t \) is the memory unit value at time \( t \). \( W_f, W_c, W_o \) are weight matrix of the forget gate, the input gate, the output gate, and the current state unit, respectively; \( b_f, b_i, b_o, b_c \)
are the offset term of the forget gate, the input gate, the output gate, the input gate, and the current state unit, respectively.

4. Rolling bearing fault diagnosis method based on WMKPCA-LSTM network

4.1. The flow chart of bearing diagnosis method based on WKPCA-LSTM network

The WMKPCA algorithm adaptively adjusts the proportion of the RBF kernel function and polynomial kernel function by introducing hyperparameters $\lambda$, which fully exploits the advantages of the two kernel functions. The LSTM neural network has the ability to extract time dimension features of the data. The bearing diagnostic algorithm based on the WMKPCA-LSTM network does not need signal processing knowledge to manually extract features and meet end-to-end fault diagnosis. The rolling bearing condition monitoring based on WMKPCA-LSTM network mainly includes data reduction and model training. The algorithm implementation process is as shown in Fig.2:

The specific implementation process of rolling bearing condition monitoring based on WMKPCA-LSTM network is as follows:

- Data preprocessing and data reduction. The acquired data is normalized to the range [0, 1] and the WMKPCA algorithm is used to reduce the dimensions of the sample data to obtain feature vectors.
- Model training. First, the feature vector is input to the LSTM network to obtain the hidden layer output. Then the hidden layer output value is input to the softmax layer to implement the classification function. Cross entropy can be used to define a loss function. Finally, the weight gradient and offset of the LSTM neural network are adjusted by the stochastic gradient descent method to minimize the error.
4.2. Softmax regression

The LSTM neural network can automatically extract the original signal features, but does not have the classification function. Therefore, a discriminative network structure is added at the top of the network. This paper adopts the softmax regression model. The softmax regression mainly implements multi-classification function \[12\], and its model loss function is as follows:

\[
J(\theta) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{K} y_{ij} \log h_{\theta}(x_i) + (1 - y_{ij}) \log (1 - \log h_{\theta}(x_i))
\]

(11)

Where \((x_i, y_i)\) is a training sample, \(K\) is the number of categories, \(N\) is the number of samples, and \(\theta\) is the model parameter.

4.3. Grid search algorithm

Hyperparameters are special parameters in machine learning that needs to be set before the model training. Although the deep neural network has good versatility and generalization, its network structure and hyperparameters settings still have a decisive influence on its performance. At present, common hyperparameter optimization methods include network search algorithm, random search algorithm, particle swarm optimization algorithm and so on. Among them, the network search algorithm is the most simple and practical and is suitable for optimizing fewer parameters \[13\]. In this paper, the network search algorithm is used to optimize the balance parameters \(\lambda\).

5. Experimental verification

5.1. Model Performance Analysis

The experimental data is from the bearing experimental data of Case Western Reserve University and the bearing model is SKF6205. First, a single point of failure was placed on the bearing using EDM technology. The fault diameters were 0.007, 0.014, 0.021, and 0.028 inches (1 inch = 0.254 mm), respectively, and data were collected by the acceleration sensors at the drive end and the fan end, respectively. In this paper, the drive end bearing data is as the experimental data, the detailed information is shown in Table 1.

The experimental data is divided into a training set and a test set, each set of samples containing 512 sample points. First, the original data is reduced to 256 dimensions by using the WMKPCA algorithm to obtain the feature vector.; Then, the feature values are input into the LSTM network with classification function in the form of \([\text{batch\_size},16,16]\)(Setting parameters are obtained by multiple tests); finally, the test set is used to verify the model. The batch\_size is 200 and the number of iterations is 100.

| Table 1. Data of the Case Western Reserve University experiment |
|---------------------------------------------------------------|
| Bearing condition    | Damage diameter (mm) | Rotating speed (r/min) | Sampling frequency (Khz) |
|----------------------|----------------------|------------------------|--------------------------|
| normal               | 0.178                | 1730                   | 42                       |
| Inner ring failure   |                      |                        |                          |
| Outer ring fault     |                      |                        |                          |
| Ball failure         |                      |                        |                          |

The loss value and accuracy of the training sample and the test sample after each iteration are recorded, and the loss value curve and the accuracy rate curve of the training data and the test data are plotted, as shown in Figure. 3 and Fig. 4. The accuracy rate is the ratio of the number of correctly classified samples to the total number of samples in each set of data.
After 100 times iteration, the loss values of the test data and the training data tend to be stable, the accuracy of training data is 98.9%, and the accuracy of test data is about 96.0%. The trend of loss and accuracy of training data and test data is synchronized. Therefore, there is no over-fitting.

5.2. Comparison of fault diagnosis methods

In order to verify the superiority of the fault diagnosis model based on the WMKPCA-LSTM network, this paper uses PCA, KPCA, MKPCA algorithm to reduce the dimensionality of the sample data and the eigenvalues are input into the same configured LSTM network for training. The experimental results are as follows:

| Method                  | Accuracy (%) |
|-------------------------|--------------|
| PCA-LSTM                | 60.00        |
| KPCA-LSTM               | 66.97        |
| MKPCA-LSTM              | 80.60        |
| WMKPCA-LSTM             | 96.00%       |

6. Conclusion

The bearing diagnosis method based on WKPCA-LSTM network makes full use of the advantages of local kernel function RBF and global kernel function Poly by introducing hyperparameter $\lambda$. The LSTM network has a memory function to further extract the dimensional characteristics of the data space and to meet the requirements of an end-to-end system to a certain extent. It is known from the experimental results that the algorithm can accurately identify the bearing state and the diagnostic accuracy is much higher than the PCA-LSTM, KPCA-LSTM and MKPCA-LSTM methods.

Acknowledgments

The work of this paper is supported by the National Natural Science Foundation of China (Grant No. 21676012) and the Fundamental Research Funds for the Central Universities (Project XK1802-4).

References

[1] Wei Z, Wang Y, He S, et al. (2017) A novel intelligent method for bearing fault diagnosis based on affinity propagation clustering and adaptive feature selection [J]. Knowledge-Based Systems, 116(C):1-12.
[2] Pennarola F, Fanfoni M, Cannata V, Bernardi B, Napolitano A. (2016) Comparison of PCA vs KPCA for physiological noise removal in resting state FMRI [J]. Physica Medica, 32.
[3] Gu Y K, Zhou X Q, Yu D P, et al. (2018) Fault diagnosis method of rolling bearing using principal component analysis and support vector machine [J]. Journal of Mechanical Science and Technology, 32(11):5079–5088.
[4] Zhang X Q, Wang X M, Wang Y. (2011) Application Research of TE Process Fault Detection Based on PCA and KPCA [J]. Process Automation Instrumentation, 32(1):8-12.

[5] Cao L J, Chua K S, Chong W K, et al. (2003) A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine[J]. Neurocomputing, 55(1-2):321-336.

[6] Xi Z, Wei-Wu Y, Zhen Ya L, et al. (2007) Process Monitoring and Fault Diagnosis of Condenser Using KPCA and PSVM[J]. Proceedings of the CSEE, 27(14):56-61.

[7] Lee J M, Yoo C K, Lee I B. (2004) Fault detection of batch processes using multiway kernel principal component analysis[J]. Computers & Chemical Engineering, 28(9):1837-1847.

[8] Zhang Y, Chi M. (2011) Fault diagnosis of nonlinear processes using multiscale KPCA and multiscale KPLS[J]. Chemical Engineering Science, 66(1):64-72.

[9] Wang L Y, Wu B, Du Z M, Jin X Q. (2018) Fault diagnosis based on short and long term memory networks in the data center air conditioning system sensors [J]. Journal of Chemical Industry and Engineering (China), 69(S2):252-259.

[10] Zimmermann M, Ghazi M M, Ekenel H K, et al. (2017) Visual Speech Recognition Using PCA Networks and LSTMs in a Tandem GMM-HMM System[J].

[11] Liu W, Wen Y, Yu Z, et al. (2016) Large-Margin Softmax Loss for Convolutional Neural Networks[J].

[12] Hochreiter S,(1997) Schmidhuber J. Long short-term memory[J]. Neural computation, 9(8):1735-1780.

[13] REUDERINK B. Fusion for audio-visual laughter detection[J]. Journal of Clinical Biochemistry & Nutrition, 2017, 40(2):141-147.