Convolutional neural network based fault detection for traction motor

Huizhong Wang[2], Linhan Qiao[2,3], Jianhai Li[2], Keke He[2]

Abstract. In propose of the feature learning for condition monitoring of inner ring and outer ring of traction motor, this paper had the method of fault detection of convolutional neural network. The goal of this approach is to autonomously learn useful features for bearing fault detection. In this method, the one-dimensional vibration signals are extracted by convolution and maximization pooling. Then the input full connection layer is used for fault classification. Compared with the traditional Back propagation neural network, Supporting vector machine with Principal component analysis, the results showed that the method in this approach has a simple structure and can achieve good results in a short period of time for medium sample data. The former achieves an accuracy of 97.71 percent on average and the latter achieves an accuracy of 88.24 percent.

Key words. Traction motor, fault detection, convolutional neural network, feature learning.

1. Introduction

Currently, traction motors are widely used. To reduce operational costs, prolong the lifetime of machines and enhance operational uptime, condition monitoring(CM) is required [1]. Convolutional Neural Network (CNN) is a method which is widely used in Feature Learning(FL). In this paper, according to the one-dimensional data of the traction motor vibration signal, convolution neural network can adopt the end-to-end feature extraction method to solve the problem of feature extraction inaccuracy and improve the accuracy of motor fault diagnosis [2]. Comparing with traditional motor fault diagnosis algorithm, it can reduce the complexity of algorithm model and the computational cost.

The remainder of this article is as follows. In the next section a literature review is given. Subsequently, the data set is discussed. Then, the feature-engineering

[2] Workshop 1 - Lanzhou University of Technology, Lanzhou, Gansu, 730050, China
[3] Corresponding author: Linhan Qiao

based approach is presented. Consequently, the feature-learning based approach is discussed. Next, the results of former system and the last are evaluated and compared. Finally, the conclusions are presented together with possible future work for the presented research.

2. Related literature

Vibration patterns depend on the machine's condition, and are therefore very suitable to detect specific conditions. For example, imbalance, which is caused by the shift between the principal axis of inertia and the axis of rotation, results in a high amplitude at the rotation frequency of the machine in frequency spectrum [3]. Other faults which can be detected

Table 1. Vibration patterns of motor condition

Motor conditions	Vibration patterns
Stator fault	$f_0, 2f_0, 4f_0, 6f_0......$
Bearing outer ring fault	$0.5Z\frac{n}{60}(1 - \frac{d}{D}\cos\beta)$
Bearing inner ring fault	$0.5Z\frac{n}{60}(1 + \frac{d}{D}\cos\beta)$

Table 2. Dimensions of the variables in table 1

Variable	Dimensions
f_0	Power frequency
Z	Number of balls or rollers
n	Speed r/min
β	Bearing pressure angle

in a similar manner are damaged raceways, since these faults generate a peak at a specific fundamental frequency [4].

To summarize, several different features with a specific goal can be extracted from vibration data. However, a human expert is still required to interpret the features to identify different machine conditions or anomalies. Machine learning is required to automate this interpretation process.

Machine learning for fault detection focuses on two major topics, i.e.: anomaly detection and fault/condition classification. Anomaly detection is process of identifying measurements that do not conform to other patterns of the data set [5].

Often, features, as discussed in the previous sub-section, are used by algorithms such as one-class Supporting vector machines (SVM), Gaussian distribution fitting, clustering in combination with principal component analysis, hidden markov models and neural networks [5-8].

In [9], the authors propose that the BP neural network nonparametric model replaces the parametric model of ARMA model to extract the structural physical parameters. However, the BP neural network has some problems such as slow learning speed and local minimum points.

Literature [10] proposed a modal parameter identification method based on continuous wavelet transform and back propagation neural network, which can accurately identify the modal parameters when the signal sampling time is very short, and has strong practicability and shows a certain Anti-noise ability. Due to the sudden change of the signal amplitude, the quality of the extended signal has some influence on the estimation of the damping ratio. From the time it takes to extract modal parameters, the lower the signal-to-noise ratio, the longer it takes to train a neural network.

the author combines fuzzy theory and neural network to form fuzzy neural network for on-line fault diagnosis, but ignores the problem caused by the cross-entropy loss function of neural network.

In this paper, fault detection model is based on convolutional neural network(CNN) . It has been proven successful in many domains . The CNN have several advantages compared to other feature-learning techniques. First, similar to stacked SAOs, CNN autonomously learn multiple levels of representations of the data through their layered structure. This enables complex features to be learned . Second, a CNN is an end-to-end learning system, therefore, only a single system has to be optimized. Finally, CNNs are used to exploit the spatial structure in the data.

A CNN works as follow: given an input containing multiple channels, such as several vibration signals combined, a CNN layer computes a similar transform as the one in Eq.(1), with the difference that the adjustable parameters of the layer are organized as a set of filters (or filter bank) and convolved over the input to produce the layer's output. The output of a CNN layer is 3D tensor, which consists of a stack of matrices called feature maps, and can be used as input to a higher level layer of the CNN model. The weights in the filter bank are shared over the input, which effectively exploits the local spatial statistics, while reducing the number of trainable parameters. The operation can be represented.

$$X_k = \sigma(W_k X_{k-1} + b_k) \tag{1}$$

$$X_k^{(m)} = \sigma(\sum_{c=1}^{C} W_k^{(c,m)} * X_{k-1}^{(c)} + B_k^{(m)}) \tag{2}$$

In Eq.(2), the layer of the network is denoted with k as before, and the $*$ operator is used for the 2D convolution of channel $c = 1, \cdots, C$ of the input X_{k-1} and the filter $W_k^{(c,m)}$, which is responsible for the $m - th$ output feature map $X_K^{(m)}$, where $m = 1, \cdots, M$. The matrix $B_k^{(m)}$ contains the bias weights. Finally, a nonlinear

activation function σ is applied to the sum of convolutions to obtain the final output.

$$F_{ij} = S(\sum_{i=1}^{n} \sum_{j=1}^{n} (M_{ij} C_{ij}) + b_1) \qquad (3)$$

Eq. (3) is the convolution calculation. Let the input layer be a matrix of convolution kernels, where n is the kernel size of the convolution kernel. filters indicates the number of output convolution layers. In practice, many convolution layers are output at a time. Strides indicates how long the convolution layer performs a scan at the interval when scanning at the input layer. Padding refers to whether the edge is scanned. If it is valid, only the known matrix is scanned. If it is same, the padding is made according to the situation to make the output equal to input-size / strides.

After convoluted, the signal will be transmitted in a pooling layer. The pooling layer usually functions to reduce the compression information and reduce the dimension of the feature mapping. However, the most important information is retained, which is equivalent to extracting the signal. The most commonly used For max-pooling, that is, "pinch" within a given area, leave the maximum value, and discard the other values. Average-pooling can also be used to leave the average.

In addition, Dropout is a strategy of reducing over-fitting. In order to prevent some special weights from occurring, each node can participate in the network update process and randomly cover part of the nodes during the neural network update process.

Fully Connected Layer is a Multi-Layer Perceptron that uses the softmax activation function as the output layer. The nodes between layers are connected in two by two. Its purpose is to classify using advanced features obtained from convolution and pooling.

3. Methodology

Our methods for motor fault detection is described in this section. In order to compare with traditional techniques discussed above, we performed experiments on a data set which was created using the test set-up below.

3.1. Test set-up

A visualization of our set-up source(the Case Western Reserve University Bearing Data Center Website) is shown in Fig.1 and the technical specifications are summarized in Table 3.

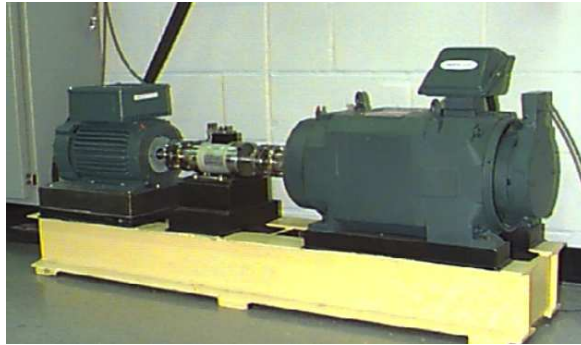Table 3. Technical specifications of the test set-up

Fig. 1. The used set-up source

Property	Value
Rotation speed	25HZ
Sample frequency	12000HZ
Accelerometer type	IEPA
Accelerometer product type	4534-B
DE	drive end accelerometer data
FE	fan end accelerometer data

The test stand consists of a 2 hp motor (left), a torque transducer/encoder (center), a dynamometer (right), and control electronics (not shown). The test bearings support the motor shaft. Vibration data was collected using accelerometers, which were attached to the housing with magnetic bases. Accelerometers were placed at the 12 o'clock position at both the drive end and fan end of the motor housing. During some experiments, an accelerometer was attached to the motor supporting base plate as well. Vibration signals were collected using a 16 channel DAT recorder, and were post processed in a Matlab environment. All data files are in Matlab (*.mat) format. Digital data was collected at 12,000 samples per second. Speed and horsepower data were collected using the torque transducer/encoder and were recorded by hand. Outer raceway faults are stationary faults, therefore placement of the fault relative to the load zone of the bearing has a direct impact on the vibration response of the motor/bearing system. In order to quantify this effect, experiments were conducted for both fan and drive end bearings with outer raceway faults located at 3 o'clock (directly in the load zone), at 6 o'clock (orthogonal to the load zone), and at 12 o'clock.

3.2. Data set

For every condition, three kinds of data set were tested, i.e., healthy bearing, outer ring fault bearing, inner ring fault bearing. The healthy bearing data set is shown in Fig.2. and the other two fault data sets are similar as the healthy one.

In general, training data set needs to be independent and identically distributed. In python environment, the data itself is subject to normal distribution and the co-variance matrix is observed. The data is also found to be independent and unrelated and therefore can be learned. Fig.3. is the healthy state of the fan side of all the vibration data. Fig.4. for the healthy state and the fault state joint probability distribution, where the sub-map (a) healthy state and the inner fault joint probability distribution, sub-map (b) for the normal state and the inner ring Fault, outer ring joint probability distribution.
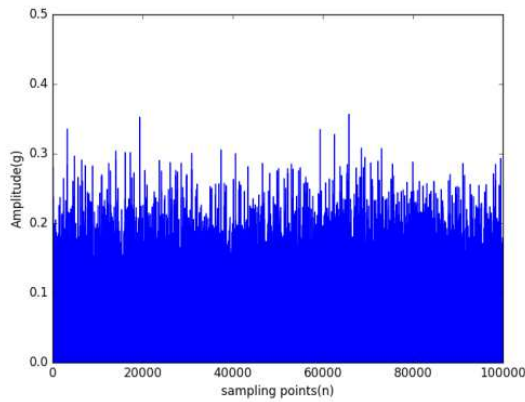

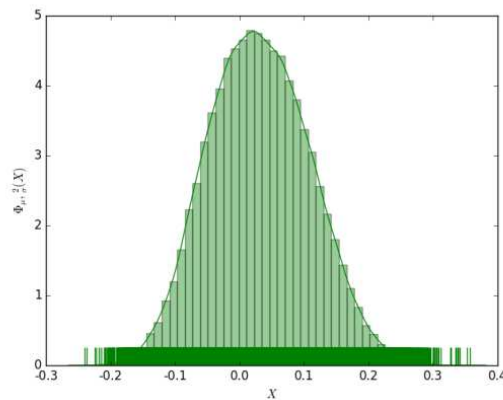
Fig. 2. Healthy bearing data set

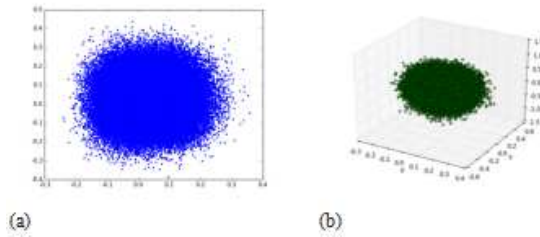Fig. 3. Fitting parametric distributions of healthy bearing data set

Fig. 4. The joint distribution of three data sets. (a) The joint distribution of
Healthy data set and the inner fault of bearing data set. (b) The joint distribution
of three kinds of data sets

It can be seen from the Fig.4. that the joint probability distribution of the data
set is elliptic or ellipsoidal, so there is a positive definite diagonal matrix that makes
the data set relevant. In the next section, the feature learning ,applied on our CNN
is discussed in detail.

3.3. K-folds cross validation

For the collected datasets, the training set and the test set are divided by the
K-folds Cross Validation (K-CV) method, as shown in Fig.5 is a schematic diagram
of a common 10-folds cross-validation algorithm.

This method firstly divides the training set and the test set, secondly, the training
set is divided into 10 parts, each of which is selected for training, and the remaining 1
part is used as the verification set so that the verification set traverses all the training
sets, and then the training result Integration, the method used here is to find the
mean. By cross-validation training neural network can learn from many directions,
to avoid falling into the saddle point. The hidden layer in CNN is relatively complex
with many saddle points. The descending direction of stochastic gradient descent
adopted is indefinite, and it is easy to fall into the saddle point. This cross-validation
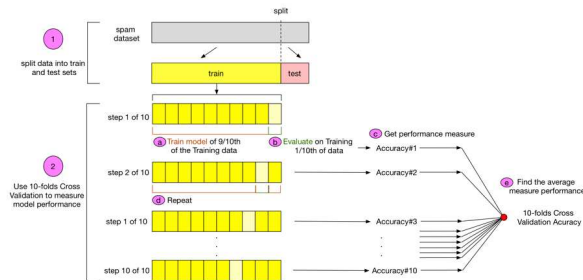by K can solve the problem.

Fig. 5. 10-folds Cross Validation

3.4. Feature learning

The feature learning based approach uses two pipeline system as depicted in Fig.6. As can be seen, the binary classification problem of balanced versus imbalanced samples were already been solved effectively using pipeline one. Therefore, we reuse this pipeline here. Nevertheless, for the detection of three or more specific bearing conditions, a feature learning model is proposed, which forms the second pipeline.
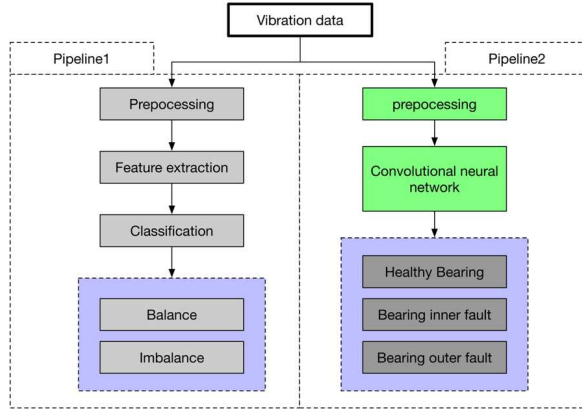

Fig. 6. High level representation of the proposed feature-learning architecture.

Our proposed feature-learning approach is based on a convolutional neural network. More specifically, a CNN model similar to the one proposed by Slavkovikj et al.The CNN model was trained using Stochastic gradient descent SGD, the differences with the mini batch gradient decent were shown in Eq.(4). And Eq. (5).

$$Dountil \; \|\theta_{new} - \theta\|_2^2 < \varepsilon \left\{ \theta \leftarrow \theta - \alpha \frac{1}{N} \sum_{i=1}^{N} \nabla_\theta L(X_i, Y_i) \right\} \qquad (4)$$

$$Dountilconverge \left\{ \theta \leftarrow \theta - \alpha \frac{1}{N_S} \sum_{i \in S} \nabla_\theta L(X_i, Y_i).S : mini - batch \right\} \qquad (5)$$

There are many advantages as using SGD. First, larger batches compute more accurate gradient estimates, but the returns are less than linear. Second, small batch processing adds noise to the learning process, so there are some regularization effects. It's helpful in jumping out of saddle points or local minima Third, as long as no repeated samples are used, it will follow the gradient of minimizing the true generalization error.

It has been shown that by using a deep architecture, i.e., a network with many layers, the network becomes more robust to the variation in the data .

4. Results

To assess the feature learning based approach, the evaluation metrics, the evaluation procedure, and the obtained results are discussed in this section.

4.1. Evaluation metrics

To quantify the performance of different classifiers, four error measurements are calculated: accuracy, precision, recall and F1-score for which the Formulas can be seen in Eqs.(6)-(9), with $|TP|$ being the amount of true positive classifications; $|TN|$, the amount of true negative classifications; $|FP|$, the amount of false positive classifications, e.g. a false alarm, and $|FN|$, the amount of false negative classifications, e.g. missed faults.

$$accuracy = \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|} \tag{6}$$

$$precision = \frac{|TP|}{|TP| + |FP|} \tag{7}$$

$$recall = \frac{|TP|}{|TP| + |FN|} \tag{8}$$

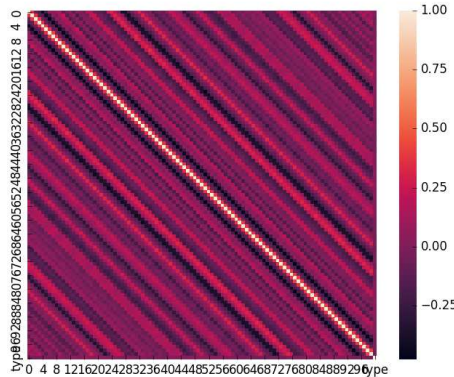$$f1 - score = 2\frac{precision * recall}{precision + recall} \tag{9}$$

Fig. 7. Data set heat map

4.2. Pipeline two results

The whole data set heat map is in Fig.7.The results achieved by pipeline two are summarized in Tale 4. The data set is dimensioned by manifold algorithm and

then input to SVM of nonlinear kernel function for classification. This method is convex optimization method and has many advantages of kernel method, but the accuracy is lower. BP algorithm updates the weights and deviations of the network by using the chain rule. Although the good classification results can be obtained, the gradient descent method itself easily leads to a local minimum. According to the multi-directional learning of the cross_validation, the direction of stochastic gradient decent and saddle point caused by multilayer neural network can be solved. Based on Tensorflow, it is also possible to quickly build a CNN model in a python environment.

Table 4. Performance results of the two-pipeline system, which uses 10-folds cross validation, executed 10 times. BPNN=BP neural network using SGD, SVM1=SVM using a linear kernel (C=10^5), SVM2=SVM using a radial basis function kernel (C=10, $\gamma = 0.3$), CNN=CNN using 10-folds cross validation and SGD

Metric	BPNN	SVM1	SVM2	CNN
Accuracy	88.24%(σ=8.05%)	72.5%(σ=18.37%)	77.5%(σ=18.37%)	97.71%(σ=6.87%)
Precision	89.83%(σ=8.20%)	73.75%(σ=19.70%)	82.08%(σ=15.78%)	95.62%(σ =6.0%)
Recall	88.24%(σ=8.05%)	72.5%(σ=18.37%)	77.5%(σ=18.37%)	97.3%(σ =6.86%)
F1-score	86.73%(σ=8.12%)	73.12%(σ=19.01%)	79.73%(σ=16.98%)	95.03%(σ=6.47%)

5. Conclusion and future work

In this article feature learning is used in the form of CNN model, which is an end-to-end machine learning system using 10-folds cross validation and stochastic gradient decent. The above analysis shows that the proposed method reduces the complexity of the traditional fault diagnosis algorithm to a certain extent and retains a good accuracy and timeliness, which provides a basis for further on-line diagnosis of motor faults. How to choose the proper method of pooling in the fault diagnosis of motor and how to get the suitable learning step by linear search method are our future concerns.

References

[1] T. ERRELL, J. ELON, W. M. NEEDELMAN, J. P. KYLE: *Wind Turbine Tribology.* Green Energy & Technology *49* (2012), 483–530.

[2] Y. LECUN, Y. BENGIO, G. HINTON: *Deep learning.* Nature *521* (2015), No. 7553, 436–444.

[3] M. MONTE, F. VERBELEN, B. VERVISCH: *The use of orbitals and full spectra to identify misalignment.* Structural health monitoring *5* (2014), 215–222.

[4] B. P. GRANEY, K. STARRY: *Rolling element bearing analysis.* Materials Evaluation *70* (2012), No. 1, 78–85.

[5] A. PURARJOM, A. H. GHAPANCHI, M. ESMALIFALAK: *A data mining approach for fault diagnosis: An application of anomaly detection algorithm.* Measurement *55* (2014), No. 3, 343–352.

[6] O. GERAMIFARD, J. X. XU, C. K. PANG: *Data-driven approaches in health condition monitoring — A comparative study.* IEEE International Conference on Control and Automation,IEEE (2010) 1618–1622.

[7] Q. MIAO, D. WANG, M. PECHT: *A probabilistic description scheme for rotating machinery health evaluation.* Journal of mechanical science & technology *24* (2010), No. 12, 2421–2430.

[8] A. VERMA, Z. ZHANG, A. KUSIAK: *Modeling and prediction of gearbox faults with data-mining algorithm.* Journal of solar energy engineering *135* (2013), No. 3, 1–11.

[9] X. U. BIN, A. S. GONG, H. E. JIA: *A novel structural parameter extraction methodology with neural network based nonparametric model.* Engineering mechanics *28* (2011), No. 4, 35–41.

[10] Y. DAI, J. X. ZHANG: *Identification of modal parameters based on wavelet transform and back propagation network.* Journal of vibration & Shock *31*, (2012), No. 3, 55–59.