Sequential Mechanical Fault Diagnosis in High Voltage Circuit Breaker using Attention Mechanism

Baoyu Zhang¹, Xinru Wang¹,*, Jin Wang²

¹ School of Information and Control Engineering, Liaoning Petrochemical University, Fushun 113001, Liaoning, P. R. China
² Minmetals Mining (Anhui) Engineering Co. Ltd
*Corresponding author email: wangxinru0908@163.com

Abstract. To address the shortcomings of traditional shallow-based model in terms of feature extraction and generalization capacity, this paper provides a high-voltage circuit breaker mechanical condition online detection scheme by using attention mechanism to locally weight the sample correlation and combining convolutional neural network (CNN) and long short-term memory (LSTM) network. The network uses convolutional layers for feature transformation of the raw vibration data, combined with the local time-domain feature representation capability of the gated unit, to extract fault-sensitive features. The temporal deep learning model using the attention mechanism enables it to extract global features of long-time sequences while making important features obtain higher weight parameters. The resulting model improves the overall learning and recognition efficiency. The experiments on the vacuum circuit breaker show that the proposed model has better performance than the classical diagnostic method based on the support vector machine.

Keywords: circuit breaker, attention mechanism, mechanical fault diagnosis.

1. Introduction

High-voltage circuit breaker is one of the most critical equipment for controlling and protecting circuits in modern power systems [1-3]. Although the circuit breakers have a long service life, the performance of the circuit breaker will be greatly degraded with the growth of service time [4-6]. Most of the failures occurring on high-voltage circuit breakers are caused by their poor mechanical properties, mainly including: aging springs, jammed mechanisms, solidified grease, broken connecting rod drive shafts, etc. Traditional periodic maintenance is time-consuming, frequent, and even reduces the reliability of the circuit breaker.

The detection signals related to the mechanical fault diagnosis of high-voltage circuit breakers include vibration signals, breaking and closing solenoid coil current and voltage, and dynamic contact displacement [3]. The transient changes in circuit breaker operation can be effectively reflected by mechanical vibration signals. The mechanical fault diagnosis of high-voltage circuit breaker based on vibration signal mainly includes the signal pre-processing, the feature extraction and the fault identification. Signal pre-processing can eliminate the zero drift of the signal and filter high frequency noise, specifically wavelet decomposition reconstruction method, five-point triple smoothing method, zero-phase digital filtering and other methods. Compared with the Fourier transform, wavelet analysis is able to provide a finer decomposition of the high frequency part.

With the increasing development of artificial intelligence technology, deep learning methods have been introduced into fault diagnosis research. By using the structure of deep layers, it has been
demonstrated that deep models have better ability to extract the fault-sensitive feature compared to shallow models [4,5]. For example, Zhou et al. [6] proposed a CNN with identity mapping module to extract feature information and identify mechanical fault. Xu and Luo [7] explored an online mechanical fault classification with a convolutional sequential deep network. The spatial and temporal features related with the fault conditions can be suitably extracted by the convolutional and recurrent layers, respectively.

In recent years, recurrent neural network (RNN) has been used in natural language processing, machine translation and time-series prediction. However, RNN suffers from gradient disappearance and would be unable to capture the long-time dependencies in the sequence. LSTM network completely considers the long- and short-term dependencies in time series. It greatly avoids the occurrence of gradient disappearance [8]. The encoder-decoder model based on the attention mechanism was originally used to solve the machine translation problem. It has been shown good performance in time series prediction due to its ability to adaptively extract the input data features while memorizing the long-term dependencies of time series [9].

In this paper, we propose a method for evaluating the condition of high-voltage circuit breakers based on the attention mechanisms. Compared with the traditional method for fault diagnosis, the continuous feature of vibration signal data can be mapped by convolutional operation. The extracted features are fed to the LSTM network with a time-step. The long-term dependency problem in the traditional RNNs can be effectively solved by the subsequent LSTM. The local temporal features related with the breaker state are preserved by the attention mechanism. Experimental results on the faults simulated on a real vacuum breaker show that the proposed method achieves the promising performance compared with the other traditional methods.

2. Deep encoder-decoder network

Assuming that the collection can be expressed as an ordered real-valued time series, $X = (X_1, X_2, \cdots, X_T)$, $X_t \in \mathbb{R}^{D_0}$, where $t \in [1, T]$ is the time-step and $D_0$ is the number of sensors. Assuming that the number of convolutional layers in the CNN is $L$, we can use 1D filter to capture fault-sensitive spatial features in the vibration data. The filter on convolution layer $l \in \{1, \cdots, L\}$ is represented as a tensor form, $A^{(l)} \in \mathbb{R}^{D_l \times d \times D_{l-1}}$, where $d$ is the duration, $D_l$ and $D_{l-1}$ are the lengths of the feature vectors on the current convolution layer and the previous layer, respectively. The feature map of the previous layer is convolved with a learnable convolution kernel, and the output after the activation function forms the neurons of this layer. The feature extraction layer is,

$$E_{i,t}^{(l)} = f \left( A^{(l)} \cdot E_{i,t}^{(l-1)} + b^{(l)} \right) = f \left( \sum_{t'=1}^{d} (A^{(l)} \cdot E_{\cdot,t'+d}^{(l-1)} + b^{(l)}) \right), \forall t \in [1, T]$$

(1)

where $E_{i,t}^{(l)} \in \mathbb{R}^{D_l}$ is a non-normalized activation on the current layer, $E_{i,t}^{(l-1)} \in \mathbb{R}^{D_{l-1 \times D_{l-1}}}$ is the normalized activation matrix on the previous layer, the symbols $\cdot$ and $\langle \cdot \rangle$ denote the convolution and vector inner product operations, respectively. $f(\cdot)$ is a nonlinear function, such as a rectified linear unit (ReLU). To efficiently compute activation values over a long cycle time, CNNs need to downsample the output of the convolutional layer using a pooling layer. The pooling is generally a max operation on the filter at the step, which is,

$$m = \max_i \{E_{i,t}^{(l)}\}$$

(2)

The pooled activations are normalized as,

$$E_{i,t}^{(l)} = \frac{1}{m + \varepsilon} E_{i,t}^{(l)}$$

(3)

where $\varepsilon$ is a small number. The max pooling operation does not only eliminate small offsets and distortions of the signal, but also has translation invariant.

To further capture the temporal correlation in the feature sequence, the spatial feature vectors extracted from the pooling layer are fed into the sequential layer. Subsequently, a bidirectional RNN model is introduced for the sequence layer. The gated recurrent unit (GRU) is placed on this model to control the forget and input gates using a gating controller.

$$g_t^2 = \sigma(W^2 h_{t-1} + U^2 E_t)$$

(4)
\[
g_t^g = \sigma(W^g h_{t-1} + U^g E_t) \\
h_t = \tanh(W^h (g_t^g h_{t-1}) + U^h E_t) \\
h_t = g_t^r h_{t-1} + (1 - g_t^r) h_t 
\]
where \(g_t^g\) and \(g_t^r\) denote update gate and reset gate, respectively; \(W^g\), \(W^r\) and \(W^h\) denote the weighting matrix of the previous moment; \(U^g\), \(U^r\) and \(U^h\) are the weighting matrix of the corresponding input vector.

3. **Attention mechanism for sequential fault diagnosis**

As the input sequence gradually increases, it is difficult for the model to retain all the information at the decoding. As a result, the performance of the model will decrease as the length of the input sequence increases. To solve this problem, we introduce an attention mechanism in the GRU decoder section that adaptively determines the relevance of the GRU encoder corresponding to the implicit state at all moments. Specifically, a vector generated from the sequence of the hidden states \(\mathbf{c}_t\), \(k = 1, \ldots, T\), at position \(k\),

\[
\mathbf{c}_k = \sum_k \beta_k^t \mathbf{h}_k 
\]

where \(\beta_k^t\) is the weight of each hidden state, which can be given as,

\[
\beta_k^t = \frac{|\mathbf{p}_k^t|}{\sum_{n=1}^M |\mathbf{p}_n^t|} 
\]

where the alignment model \(p_k^t\) is learned by the following equation,

\[
p_k^t = \sigma(W_a \mathbf{h}_k \mathbf{x}_t) \\
\mathbf{h}_{t,k} = \tanh(\mathbf{x}_t^\top \mathbf{W}_k + \mathbf{x}_t^\top \mathbf{W}_k) 
\]

where \(W_a\) and \(W_t\) are learnable weights. The structure of the model is shown in figure 1.

![Figure 1. Structure of attentional mechanism](image)

4. **Case study on vacuum circuit breaker**

In this section, the performance of the proposed method is evaluated on the ZW32-12FG/630-20 vacuum circuit breaker. The actual measured vibration signal is converted into the corresponding voltage using a YD-39 acceleration sensor. The measurement range of the sensor is 0-5500 m/s, and the sampling rate is 10kHZ. To avoid additional interference signals, the sensors are placed in the vacuum circuit breaker's shell to collect the vertical vibration signal. The NI9234 data acquisition card is responsible for the part that converts the voltage pairs into digital data output. Three vibration signal types were used in the test experiment: normal condition; decoupling closure solenoid blocked; spindle blocked.
4.1. Parameter configuration and evaluation metrics

In the attention deep network model, the designed CNN has 1 convolutional layer and 1 pooling layer, and the number of convolutional filters is 64. The size of the convolution kernel is set to 1 and the step size is 1. The size of MaxPooling1D in the pooling layer is 1. ReLU function is used for the activation function in the CNN network. The number of network layer nodes used in the LSTM model is 100, and the activation function is a sigmoid function. The experiment uses the method of comparing the SVM model with the CNN-LSTM model. Samples of the same mechanical vibration signals are input for fault diagnosis, and the diagnostic results of both models are evaluated in ROC and PRC curves.

4.2. Fault classification and performance analysis

For the clogging status of decoupling closure solenoid, the ROC curve using the deep network model of the attention mechanism completely envelops the ROC curve of the normal state. It means that the fault signal can be accurately classified with the normal signal, as shown in figure 2. The curves for the normal state and the decoupling closure solenoid blockage fault state in figure 2 (a) keep changing smoothly with threshold changes, while the normal state curve using SVM in figure 2 (b) shows large fluctuating changes during this period. From figure 2 (a) and (b), it can be seen that the values of AUC for both normal state and decoupling closure solenoid blockage faults are 0.99 for the attention mechanism, while the corresponding AUC values are 0.52 and 0.52 for the SVM model. The AUC values of the deep network model with attention mechanism are much larger than those of the SVM model. In summary, it can be seen from the ROC curves that the attention mechanism model has a better diagnosis effect on the state of the mechanical vibration signal of the circuit breaker.

Precision-recall curve (PRC) can be used to further determine the classification performance of the model if the ratio of normal state signal samples to fault state signal samples is disparate. The PRC curve using the attention mechanism is closer to the upper right corner of the coordinate axis compared to the SVM-based PRC curve, as shown in figure 3. The normal state and the fault state curve of the occurrence of decoupling closure solenoid blockage in figure 3 (a) always remain smooth with the change of threshold value during the diagnosis process, and the detection rate remains smoothly changing in the process of increasing the detection rate. The values of AUC for the normal state curve and the circuit breaker blockage fault curve with disconnected closing solenoid using the attention mechanism network model are 0.996 and 0.977, respectively, while the values of AUC corresponding to the SVM model are 0.575 and 0.629. It can be concluded that in the PRC curve, the attention mechanism network has higher sensitivity and accuracy in diagnosing the status of the circuit breaker both during normal operation and when blockage of the decoupling closure solenoid occurs.

![Figure 2](image1.png)

(a)

![Figure 2](image2.png)

(b)

Figure 2. ROC curves of (a) attentional mechanism network and (b) SVM over the decoupling closure solenoid blockage fault mode. The micro- and macro-averaging ROC curves are denoted by dashed pink and navy-blue lines, respectively.
Figure 3. P-R curves of (a) attentional mechanism network and (b) SVM over the decoupling closure solenoid blockage fault mode. The micro- and macro-averaging P-R curves are denoted by dashed pink and navy-blue lines, respectively. The horizontal line (red with dashes) represents the random performance level.

For the spindle blockage condition, the experimental results of using the attentional mechanism network model are shown in figure 4 as well as in figure 5. The attentional mechanism network model has good performance in the degree of ROC curve close to the upper left corner, as well as the smoothness of the model ROC curve and the magnitude of change with threshold. The macro-averaging AUC of the attentional mechanism network model is 0.99, which is 0.45 higher than that of the SVM model. It is concluded that the attentional mechanism has excellent performance in the diagnosis of vibration signals in the normal and spindle blockage states of the circuit breaker.

Figure 4. ROC curves of (a) attentional mechanism network and (b) SVM over the spindle blockage fault mode. The micro- and macro-averaging ROC curves are denoted by dashed pink and navy-blue lines, respectively.

Figure 5. ROC curves of (a) attentional mechanism network and (b) SVM over the spindle blockage fault mode. The micro- and macro-averaging ROC curves are denoted by dashed pink and navy-blue lines, respectively.
5. Conclusion
The proposed deep learning model, which is based on the attention mechanism and the CNN_LSTM model, is introduced to improve the overall learning and recognition efficiency of the model by extracting the global features of long time series and obtaining higher feature weight parameters of important data features through the attention mechanism. The performance of the model is evaluated by using ROC curve and PR curve. For the fault diagnosis on the high voltage circuit breaker, comparing with the existing model shows that real-time monitoring using the attention mechanism receives a better detection capability on the decoupling closure solenoid blockage and the spindle blockage.

6. Acknowledgments
This work was supported by National Natural Science Foundation of China (Grants No. 61703191) and talent scientific research fund of LSHU (No. 2017XJJ-012).

7. References
[1] Cao Y, Luo L, Wang Q and Zhang J 2021. Fault Diagnosis of High-Voltage Vacuum Circuit Breaker with Convolutional Deep Network, Power System Protection and Control, 49(3), 39-47.
[2] Zhang Y, Jiang Y, Chen Y and Zhang Y 2020. Fault diagnosis of high voltage circuit breaker based on multi-classification relevance vector machine, Journal of Electrical Engineering & Technology, 15, 413-420.
[3] A. Asghar Razi-Kazemi and K. Niayesh 2021. Condition monitoring of high voltage circuit breakers: past to future. IEEE Transactions on Power Delivery, 36(2), 740 – 750.
[4] Luo L, Xie L and Su H 2020. Deep learning with tensor factorization layers for sequential fault diagnosis and industrial process monitoring, IEEE Access, 8, 105494-105506.
[5] Wang Q and Luo L 2020. Sequential fault detection and classification in chemical process using a deep convolutional encoder-decoder architecture, IOP Conference Series: Earth and Environmental Science, 696, 012052.
[6] Zhou K, Lin X, Wang X, Zhou W and Li W 2020. A CNN Based Approach with identity mapping module for mechanical fault diagnosis of high voltage circuit breaker. Journal of Physics: Conference Series, 1601, 062048.
[7] Xu D and Luo L 2020. Online fault diagnosis of high-voltage vacuum circuit breaker based on deep convolutional long short-term memory network. Proc. Int. Conf. on Robotics and Rehabilitation Intelligence vol 1335(Fushun: China/Springer), p 137-149.
[8] Chen S, Luo L, Xia Q and Wang L 2020. Self-attention mechanism based dynamic fault diagnosis and classification for chemical processes, Journal of Physics: Conference Series, 1914, 012046.
[9] Mu K, Luo L, Wang Q and Mao F 2021. Industrial process monitoring and fault diagnosis based on temporal attention augmented deep network. Journal of Information Processing Systems, 17(2), 242-252.