Research on Feature Extraction Method of Converter Transformer Vibration Signal Based on Markov Transition Field

Pengfei Wang¹, Gang Yu¹, Huafeng Wu¹, Zhanlong Zhang², Rui Xiao²

¹ Dali Bureau of EHV Power Transmission Company, China Southern Power Grid
² State Key Laboratory of Transmission and Distribution Equipment and System Safety and New Technology, Chongqing University

*Corresponding author’s e-mail: 201911131082@cqu.edu.cn

Abstract. In view of the high complexity of the vibration signal of the converter transformer and the large amount of data, the construction of the feature extraction model of the converter transformer based on the vibration signal is difficult and the accuracy is not high. This paper proposes a feature extraction model of commutation vibration signals based on Markov transition field and residual convolutional neural network. This paper first divides the interval according to the signal amplitude and calculates its Markov transition field matrix, and then obtains the two-dimensional representation of the one-dimensional vibration signal by calculating the Markov transition field matrix through the transition matrix. Finally, the feature extraction of the vibration map is performed through the residual convolutional neural network. Analysis of the actual measured data at the converter station shows that the average working condition recognition accuracy of the model in this paper reaches 93.1%. It is better than classic time series processing networks such as long and short-term memory networks and one-dimensional convolutional neural networks. It solves the problem of difficulty in building and training long vector deep learning networks by constructing two-dimensional representations of one-dimensional vectors. It is based on commutation The research on fault detection methods of variable vibration signals provides the basis.

1. Introduction

The converter transformer is an important equipment of the DC transmission system, and its safe operation is directly related to the safe operation of the DC transmission system. At present, the method of evaluating the operating status of transformers based on vibration signals has received extensive attention. This method uses the transformer vibration signal data collected by the sensor to monitor the operating status of the transformer winding and iron core. The surface vibration signal of the transformer is not only closely related to the working state of its own mechanical components (such as iron cores, windings and other external structures), its vibration signal itself also contains a large amount of rich state information. The withstand voltage level of the converter transformer is high, the physical field coupling is complicated, the insulation result is complicated, the fuel tank space and core section are larger than conventional power transformers, and the outflow of the valve hall has higher harmonics, which all lead to the vibration and noise of the converter transformer operation. Larger than power transformers, it contains more information. Therefore, how to quickly and
effectively evaluate the operation status of the converter transformer through the vibration signal is the main concern of the current power sector, and it is also an urgent problem to be solved.

For the processing of one-dimensional vibration sequences, most of the traditional methods are to manually construct mathematical models in order to obtain signal indicators that characterize the mechanical structure of the transformer to reflect changes in the operating state of the transformer. Literature [1] proposed using the instantaneous root mean square (RMS) value calculated by Hilbert transform as a fault indicator and proved that it is sensitive to the severity of the fault. Literature [2] proposed the method of ground state energy ratio to reflect the degree of transformer DC bias. Literature [3] proposed a method for diagnosing the mechanical state of transformer windings based on the combination of short-circuit reactance and mechanical vibration. Literature [4] proposed a signal feature extraction method based on empirical wavelet transform and multi-scale entropy. The feature vector of transformer vibration signal was constructed by calculating the correlation between empirical wavelet classification and the original signal. Traditional methods have good results for simple vibration signal extraction, but there are still problems: 1) The processing effect of high complexity, non-periodic and non-stationary signals is not good; 2) The calculation is cumbersome and relies on certain prior knowledge, which is not conducive to big data processing.

In recent years, with the continuous growth of computer computing power, deep learning has achieved considerable development, especially in the fields of speech recognition, image recognition, and autonomous driving. A large number of vibration signal analysis models based on deep learning have been proposed. Literature [5] proposed a self-powered RFID sensor tag, deep belief network and multi-core support vector machine transformer fault diagnosis method. Literature [6] proposed a transformer fault diagnosis method based on IoT monitoring system and integrated machine learning. An integrated machine learning system composed of a deep belief network, stacked noise reduction autoencoders with different activation functions and related vector machines is proposed. Literature [7] proposed a novel modeling framework, which includes multiple stacked sparse autoencoders and a nonlinear regression function for tool wear prediction. This model can learn more from the different feature domains of the vibration signal. Multiple features. However, it is more difficult to build a deep learning model for one-dimensional sequences. The reason is that training for long vectors is difficult and inefficient. It is easy to lose time correlation during the training process, resulting in information loss and reduced recognition accuracy.

In order to solve the above problems, this paper proposes a feature extraction model of commutation variable vibration signal based on Markov transition field and residual neural network. Firstly, the Markov transition probability matrix is calculated by taking the amplitude of the vibration signal as the interval, and then the calculation method of the Markov transition field matrix is defined, and the one-dimensional vibration signal is converted into a two-dimensional map representation through pseudo-color conversion. The problem of difficult training of long vectors is solved by converting one-dimensional signals into two-dimensional maps. After that, the residual convolutional neural network is used to further feature extraction and classification of the pictures. The effectiveness of the method is proved by the field-measured converter transformer vibration data set. The accuracy of the method is higher than that of the classical time series processing model. It can provide a basis for the research of the fault diagnosis method of converter transformer based on vibration signal.

2. Markov transition field principle

2.1. State transition probability matrix
People usually call the state of something at a certain time as the state of the thing at this time, and record the i-th state of the thing at time \( t \) as \( S_i \). Assuming that things have \( n \) different states, the state space formed by these \( n \) different states is denoted as \( S = \{S_1, S_2, \ldots, S_n\} \).
The object under study is in a certain state in the state space at time $t$, and the possibility that it is in various states is the state probability. Let the probability of state $S_i$ be $\pi_i(t)$, as shown in formula 1:

$$\pi_i(t) = p\{X_i = S_i\}, t = 1,2,...,n$$  \hspace{1cm} (1)

Let $\Pi_k$ denote the state probability space of the studied object in the $k$th period, as shown in formula 2:

$$\Pi_k = (\pi_i(k),\pi_j(k),...,\pi_n(k)), k = 1,2,...$$  \hspace{1cm} (2)

According to the principle of probability theory: $\sum_{i=1}^{n} \pi_i(k) = 1, \pi_i(k) \geq 0$

If we regard the research object as a system, then the state of the system changes over time. State transition is the transition of the system from a state $S_i$ in a period to a possible state $S_j$ in the future. The possibility of this state transition is the transition probability. The transition probability can be divided into one transition and multiple transitions. The so-called one transition refers to the state transition of the system in two adjacent periods, and the multiple transition refers to the state transition of the system after multiple periods.

The system is in state $S_i$ at time $t = m$, and the probability of being in state $S_j$ at the next time $t = m + 1$ is denoted as $p_{ij}$, then $p_{ij}$ is called the probability of a transition. As shown in formula 3:

$$p_{ij} = p\{S_j | S_i\} = p\{X_{m+1} = S_j | X_m = S_i\}$$  \hspace{1cm} (3)

The system is in state $S_i$ at time $t = m$. After $k$ transitions, the probability of being in state $S_j$ at time $t = m + 1$ is the probability of $k$ transitions, denoted as $p_{ij}$, as shown in Equation 4:

$$p\{S_j | S_i\} = p\{X_{m+k} = S_j | X_m = S_i\}, k = 1,2,...$$  \hspace{1cm} (4)

Since the probability is non-negative, and the process starts from a state, after one transition, it must reach a certain state in the state space. Therefore, the one-step transition probability satisfies:

$$\sum_{j=1}^{n} p_{ij} = 1(i, j = 1,2,...,n), 0 < p_{ij} < 1(i, j = 1,2,...,n)$$  \hspace{1cm} (5)

The matrix composed of the set of transition probability of the system is called the transition probability matrix, as shown in formula 6:

$$(p_{ij})_{n \times n} = \begin{bmatrix}
    p_{11} & p_{21} & \cdots & p_{n1} \\
    p_{12} & p_{22} & \cdots & p_{n2} \\
    \vdots & \vdots & \ddots & \vdots \\
    p_{1n} & p_{2n} & \cdots & p_{nn}
\end{bmatrix}$$  \hspace{1cm} (6)

2.2. Markov Transition Field

For a given vibration time sequence $X = \{x_1, x_2, ..., x_n\}$, the sequence is divided into $Q$ quantile array distances according to the vibration amplitude, and each $x_i$ in the sequence is mapped to a corresponding $q_i$, which can establish a $Q \times Q$-sized Markov transition The matrix $W$, $w_{ij}$
represents the probability of the element in $q_i$ being transferred to $q_j$. Obviously, the element in the matrix satisfies $\sum_{j=1}^{Q} w_{ij} = 1$.

$$W = \begin{bmatrix}
    w_{11} & w_{12} & \cdots & w_{1Q} \\
    w_{21} & w_{22} & \cdots & w_{2Q} \\
    \vdots & \vdots & \ddots & \vdots \\
    w_{Q1} & w_{Q2} & \cdots & w_{QQ}
\end{bmatrix}$$

(7)

$$w_{ij} = p\{x_i \in q_j | x \in q_i\}$$

(8)

This defines the Markov transition field $M$, $m_{ij}$ represents the transition probability of $q_i \rightarrow q_j$, that is, by considering the time position, the transition probability matrix $W$ containing the amplitude axis is extended to the $M$ matrix containing the time axis (time stamp). Finally, the corresponding gray values of the elements in the $M$ matrix are matched to a point in the color space, and the one-dimensional time series are visualized in the form of a two-dimensional map.

The overall map construction process is shown in Figure 1:

![Figure 1. MTF map generation flow chart](image)
3. Construction of residual convolutional neural network model

3.1. Principle of residual convolutional neural network

The residual convolutional neural network generally includes an input layer, a convolutional layer, a fast track network, a pooling layer and a fully connected layer. Its essence is a multi-layer supervised learning neural network. The input data is in the form of a two-dimensional matrix. The convolutional layer and the pooling layer are the core modules to realize the feature extraction function. The fast track network is a cross-layer connection model introduced to solve the problem of gradient explosion or disappearance. The cross-connection model allows each layer to be connected to non-adjacent layers, which can receive input from any layer in front, and pass its output to any non-adjacent layer in the back. The network model adopts the gradient descent method to minimize the loss function to reversely adjust the weight parameters in the network layer by layer, and improves the accuracy of the network through frequent iterative training. The low hidden layer of the residual convolutional neural network is composed of a convolutional layer, a pooling layer and a fast track network in parallel, and the upper layer is a fully connected layer corresponding to the hidden layer and logistic regression classifier of the traditional multi-layer perceptron. The input of the first fully connected layer is the feature image obtained by the feature extraction of the low hidden layer, and the output layer of the last layer is a classifier used to output the probability distribution of the result. It can use logistic regression, Softmax regression, support vector machine and other methods.

3.1.1. Principle of Convolutional Layer

The operation process of the convolution layer is to use the convolution kernel to slide on the original input according to the specified step size, perform the convolution operation, and obtain the feature map after the activation function. Then the two-dimensional convolution operation $C = A \ast B$ is defined as formula 10:

$$c_y = \sum_{s=1}^{m} \sum_{t=1}^{n} a_{i+m-s,j+n-t} \cdot b_m, 1 \leq i \leq M - m + 1, 1 \leq j \leq N - n + 1$$

In formula 10, assuming that $A$ and $B$ are matrices, the sizes are $M \times N$ and $m \times n$, respectively.

![Figure 2. Principle of Convolutional Layer](image)
3.1.2. Principle of Pooling Layer

The pooling layer reduces the amount of data processing while retaining useful information by down-sampling the feature map. Because after a certain input feature is obtained through the convolutional layer, the absolute position of the feature is not important. Only the relative position of this feature value and other feature values can respond to changes in similar objects caused by deformation and distortion. Taking maximum downsampling as an example, it is defined as Equation 11, 12:

\[ G^A_{\lambda, \tau}(i, j) = (a_{\mu})_{\lambda \times \tau} \]
\[ \max \text{down}(G^A_{\lambda, \tau}(i, j)) = \max\{a_{\mu}, (i-1)\cdot \lambda + 1 \leq s \leq i \cdot \lambda, (j-1)\cdot \tau + 1 \leq t \leq j \cdot \tau\} \]

![Figure 3. Principle of Max pooling](image)

3.1.3. Principle of Fast Track Network

In a fast track network module, the cross-layer connection generally only spans 2~3 layers, but it does not exclude more layers. The specific schematic diagram is shown in Figure 4. The calculation results \( F(x) \) and \( H(x) \) after adding the cross-layer connection (X) The following relationship exists between:

\[ H(x) = F(x) + x \]

This means that the cross-connect module calculates the residual when it is not connected. From the overall function point of view, if \( \{W_i\} \) is used to represent the ownership value of the spanning module, then the residual module actually calculates the following output result:

\[ y = F(x, \{W_i\}) + x \]

Among them, \( F(x, \{W_i\}) \) is also called residual mapping, which can be learned through back propagation. In the case of two weight layers, the calculation process can be expressed as:

\[ F(x, \{W_i\}) = w_2 \sigma(w_1 x) = w_2 \text{ReLU}(w_1 x) \]

The calculation of the residual module requires that \( F(x, \{W_i\}) \) and \( x \) have the same dimension. If their dimensions are different, an additional weight matrix \( w_i \) can be added to linearly project \( x \) to make the dimensions the same. The corresponding calculation process is:
\[ y = F(x, \{W_i\}) + w_i x \]  \hspace{1cm} (16)

3.2. Construction of feature extraction model

The following figure shows the entire two-dimensional map feature extraction network model based on the Pytorch library. The parameters of each layer of convolution kernel have been marked in the figure. The output of each layer is normalized and processed by Relu activation function to accelerate the convergence speed of network training. The fully connected layer uses dropout operations to prevent overfitting. The entire training network includes an independent convolutional layer conv1, a maximum pooling layer maxpool, a convolution residual cross-connection module, an average pooling layer avgpool and a fully connected layer.
4. Experimental analysis

4.1. Data set construction
The data comes from the routine test of the ZZDPFZ-480000/500-400 flexible DC transformer at a converter station of China Southern Power Grid. The test types are shown in the table.

| Test name               | Remarks            |
|-------------------------|--------------------|
| No-load test            | 1.0Un, 112.7Kv, 126Kv, 140.9Kv, 150Kv, 152Kv, 154.9Kv, 155Kv, 154Kv |
| Load test               | 1130A, 1470A, 1910A, 1920A |
| Temperature rise test   | 1.05pu             |

Vibration measurement adopts DH5902N data acquisition and analysis system. The sensor adopts IEPE piezoelectric acceleration sensor, model 1A941E, axial sensitivity is 100Mv/g, sampling frequency is 20kHz, and the sensor is directly attached to the surface of the converter transformer case for measurement. Select 12 measuring points, the layout of measuring points is as follows:

![Figure 6. Layout of measuring points](image)

Each operating condition corresponds to a data label, and each data label contains a Markov transition field map of the corresponding waveform. Some data sets are shown in the figure below, and the image data resolution is uniformly set to $640 \times 480$. 
4.2. Effect verification

The no-load test data under the rated voltage of 1.0Un is used as the no-load training station, and the test data under the load temperature rise of 1.05 times the quota is input into the network as the load training set. The sequence of pictures is randomly shuffled during input to prevent the neural network from being inclined to the data sequence. After the training is completed, input the remaining no-load test data of different voltage and current levels as the verification set into the trained network to obtain the classification results as shown in the figure, and the comparison results with the traditional one-dimensional time series processing network are shown in the table:
Figure 8. Comparison of accuracy of different models

| Model name  | Accuracy |
|-------------|----------|
| DWT&KNN     | 86.2%    |
| 1D-CNN      | 90.1%    |
| FCN         | 86.3%    |
| ResNet      | 90.1%    |
| LSTM        | 87.3%    |
| Paper model | 93.1%    |

It can be seen from the above figure that the proposed model in this paper has a good classification effect on the field measured data set, with an average recognition accuracy rate of 93.1%. Compared with the traditional one-dimensional time series processing model, the recognition accuracy of this model is improved by about 5.1%.

5. Conclusion

This paper proposes a converter transformer vibration signal feature extraction model based on Markov transition field and residual convolutional neural network, and constructs a two-dimensional vibration map data set based on the measured data of the converter station. The conclusions are as follows:

1) By calculating the Markov probability transition matrix and Markov transition field matrix of the time series sequence, the one-dimensional time series sequence is converted into a two-dimensional map, which makes full use of the current advantages of deep learning in computer vision and overcomes one-dimensional The sequence is difficult to train and the efficiency is not high.

2) The residual convolutional neural network is used to extract and classify the features of the two-dimensional map, and the measured data set is used for verification. The experiment proves that the average recognition accuracy of the model in this paper is 93.1%, and the classification effect of the converter transformer working condition is good. At the same time, the average recognition accuracy of the model in this paper is 5.1% higher than that of the classic one-dimensional time series processing model. Has a big improvement. It provides a method basis for fault detection and identification based on the vibration signal of the converter transformer.

References

[1] Jose R, David G, Juan P, David C and Martin V. (2020) Vibration Signal Processing-Based Detection of Short-Circuited Turns in Transformers:A Nonlinear Mode Decomposition Approach. Mathematics. pp. 8: 575–592.
[2] Xingmou L, Jiaxin W, Haibo L, Ammad J, Jing W. (2019) Research on DC bias analysis for transformer based on vibration Hilbert Huang transform and ground-state energy ratio method. pp.109: 73-82.

[3] Zhang B, Zhao D, Wang F, Shi K, Zhao Z. (2018) Research on mechanical fault diagnosis method of power transformer wingding. pp. 16: 2096–2101.

[4] Miaoying Z, Gang X. (2017) Feature extraction of power transformer vibration signals based on empirical wavelet transform and multiscale entropy. pp. 12: 63–71.

[5] Chaolong Z, Yigang H, Shanhe J, Tao W, Lifen Y, Bing L. (2018) Transformer fault diagnosis method based on self-powered RFID sensor tag, DBN and MKSVM. pp. 19: 8202–8214.

[6] Chaolong Z, Yiang H, Bolun D, Lifen Y, Bing L, Shanhe J. (2020) Transformer fault diagnosis method using IoT based monitoring system and ensemble machine learning. pp. 12: 202–224.

[7] Shi C, Luo B, He S, Li K, Liu H, Li B. (2019) Tool Wear Prediction Via Multi-Dimensional Stacked Sparse Autoencoders with Feature Fusion. pp. 8: 5150–5159.