Research on Blind Source Separation of Transformer Vibration Signal Based on Full Convolution Time Domain Audio Separation Network

Tao Li¹, *, Chunlin Guo¹, Jie Xie¹, Wenkai Li¹, Huiyuan Ma², Xiaozhong Wu³

¹ School of Electrical and Electronic Engineering, North China Electric Power University, Beijing, China
² State Grid Beijing Electric Power Company, Beijing, China
³ State Grid Hunan Electric Power Company Limited, Changsha, China

*Corresponding author’s e-mail: lit@ncepu.edu.cn

Abstract. The vibration signal on the surface of the transformer tank is formed by mixing the winding and the iron core. Blind source separation of vibration signals can diagnose the operating status of windings and iron cores separately, and improve the accuracy of using vibration signals to diagnose transformers. However, due to the high correlation between the vibration signals of the winding and the iron core, traditional blind source separation algorithms such as FastICA based on the assumption of signal independence have certain limitations. In this paper, a fully convolutional time-domain audio separation network (ConvTasNet) based on a deep learning model is used to blind source separation of transformer vibration signals. The training set and verification set are used to iterate the network. The training is completed when the network loss value is iterated to the specified requirements. This will realize the separation of the vibration signal of the winding and the iron core. This paper has carried out simulation experiments to verify the algorithm. Compared with the traditional FastICA method, the separation effect has been improved. It can more accurately separate the vibration signals of different sources, and the fault diagnosis of transformer windings and iron cores based on vibration signals Promotion and application are of great significance.

1. Introduction

The power transformer will produce mechanical vibration during operation. When the equipment fails, its vibration will also change accordingly. Through the change of vibration characteristics, it can be judged whether the equipment is in an abnormal operation state[1][2]. The vibration of the transformer is mainly produced by the electromagnetic force and magnetostriction of the winding and the iron core. The vibration signal is transmitted to the surface of the oil tank through the rigid body connector and insulating oil inside the transformer, and a mixed vibration signal is formed by superposition. The blind source separation algorithm refers to extracting and separating each source signal that cannot be directly observed from several observed mixed signals. In transformer faults, winding faults accounted for 46.4%, and iron core faults accounted for the third place among all transformer faults. Collecting the mixed vibration signal on the surface of the transformer tank, applying the blind source separation algorithm to it, and separating the winding vibration signal and the iron core vibration signal, are of great significance to the condition monitoring of the transformer winding and the iron core. In transformer monitoring, it is very necessary and
important to detect and evaluate the status of windings and iron cores in order to find hidden dangers in time and extend the life of the transformer.

2. Vibration mechanism of oil-immersed power transformer

Vibration sources of transformers in operation include windings, iron cores, and cooling components. The cooling part of the high-voltage transformer is far away from the main body of the transformer, the amplitude contribution is small, and the vibration frequency is mainly concentrated at about 20Hz, which can be clearly distinguished from other sources of vibration, so it will not be discussed again. The interaction of the current in the winding and the surrounding magnetic field causes the winding to vibrate. The vibration of the iron core mainly includes magnetostrictive effect and nonlinear magnetic effect. The vibration of the winding and the iron core is transmitted to the surface of the box body through the rigid body connection of the transformer, insulating oil and solid support rods.

The main vibration sources of the transformer are windings and iron cores. These two vibration sources are highly correlated and have overlapping frequency components. In order to better monitor the state of the transformer through the vibration signal, the blind source separation algorithm is used to estimate the winding vibration signal and the iron core vibration signal respectively, and the state of the transformer winding and the iron core can be monitored. Based on this, this paper applies the blind source separation algorithm based on ConvTasNet.

3. Research on blind source separation technology based on ConvTasNet

3.1. Overall framework

Convolutional time domain audio separation network (Conv-TasNet) consists of three processing stages\cite{5}, as shown in Figure 1: encoder, separator and decoder. First, the encoder module converts the segments of the mixed signal into corresponding representations in the intermediate feature space. This representation is then used to estimate the multiplication function (mask) of each source through the separation module. Finally, the decoder module reconstructs the source waveform by transforming the masked encoder features\cite{4}.

3.2. Page Numbers The structure of each part of the network

3.2.1. Convolutional encoder/decoder. The input mixed signal can be divided into overlapping segments. The input mixed signal can be divided into overlapping segments $x_k \in \mathbb{R}^{T_k}$ of length $L$, $k=1,...,\tilde{T}$, representing the index of the segment, and representing the total number of input segments. Convert $x_k$ to $N$-dimensional representation of $w \in \mathbb{R}^{T \times N}$ through a one-dimensional convolution operation:

$$w = \sigma(xU^T)$$

Among them, $U \in \mathbb{R}^{N \times T}$ contains $N$ vectors (encoder basis functions), and each vector has a length of $L$. $\sigma(\cdot)$ is an optional non-linear function. In this model, the ReLU function is selected to ensure non-negativeness. The ReLU function can be expressed as:

$$ReLU = \begin{cases} x & \text{if } x>0 \\ 0 & \text{if } x>0 \end{cases}$$
The decoder uses a one-dimensional transposed convolution operation to reconstruct the waveform from this representation, which can be expressed by matrix multiplication as:

$$\hat{x} = w^T \mathbf{V}$$  \hspace{1cm} (3)

Where $\hat{x} \in R^{1 \times L}$ is the estimated $x$, $V \in R^{n \times L}$ is the basis function of the behavioral decoder, and the length is $L$. Adding the overlapped reconstruction segments together is the final waveform of each source.

### 3.2.2. Estimated separation mask.

The separation of each frame is achieved by estimating $C$ vectors (masks), $m_i \in \mathbb{R}^{1 \times T}$, $i=1,...,C$, where $C$ is the number of source signals in the mixed signal, and $m_i \in [0,1]$. The $w$ obtained by applying $m_i$ to the encoder can be expressed as:

$$d_i = w \odot m_i$$  \hspace{1cm} (4)

Among them, $\odot$ means that the corresponding points are multiplied. The estimated waveform signal $\hat{x}$, $i=1,...,C$ of each source can be reconstructed by the decoder

$$\hat{x} = d^T \mathbf{V}$$  \hspace{1cm} (5)

### 3.2.3. Convolutional separation module.

The convolutional separation module is shown in Figure 2, which uses a full convolutional separation module composed of a stack of one-dimensional expansion convolutional blocks based on a time-domain convolutional network (TCN). Each layer in the TCN is composed of a one-dimensional convolution block with a gradually increasing expansion factor. In order to take advantage of the timing dependence of the vibration signal, the expansion factor increases exponentially to ensure that a sufficiently long time window is included. In Conv-Tasnet, M convolutional blocks with expansion factors of 1, 2, 4,..., $2^{M-1}$ are repeated $R$ times, and the input of each block is zero-filled to ensure that the output length is the same as the input. The output of TCN is sent to a convolution block with a convolution kernel size of 1, and then passes through a nonlinear activation function to obtain $C$ mask vectors estimated by $C$ vibration sources.

![Figure 2. ConvTasNet flowchart.](image)

### 4. Experimental Program

#### 4.1. Data set settings

In order to verify the separation effect of the ConvTasNet network in the blind source separation of transformers, this paper uses a simulation data set for experimental verification. The frequency components of the two source signals overlap, so as to test the two strong correlations similar to the transformer winding and iron core vibration signals. Source signal separation effect. The time length of each sample of the data set is 0.5s, and the sampling frequency is 15625 Hz.
The simulation signal data set is 60,000 groups of simulation signals, of which 40,000 groups are the training set, and 10,000 groups are the verification set and the test set. Each group of signals includes mixed signal, simulated source signal 1, simulated source signal 2. The two simulated source signals are set to partially overlap in frequency components, and the amplitude is randomly distributed in [-0.5,0.5].

4.2. Experimental parameter configuration
The network sets 100 epochs during training, and the initial learning rate is 0.001. If the accuracy of the validation set does not improve within 3 consecutive epochs, the learning rate is halved. The optimizer uses Adam. Table 1 shows the hyperparameters of the network.

Table 1. Network hyperparameter configuration.

| Parameter | Description | Numerical Value |
|-----------|-------------|-----------------|
| N         | Number of filters in encoder/decoder | 256 |
| L         | Filter length | 16 |
| B         | Number of 1×1 convolutional blocks in the bottleneck channel and the remaining path | 128 |
| Sc        | Skip the number of 1×1 convolutional blocks in the connection path | 128 |
| H         | Convolution channel number | 256 |
| P         | Convolution kernel size in convolution block | 3 |
| X         | Number of convolution blocks in each repetition | 8 |
| R         | Convolution block repetition times | 3 |

4.3. Training goal
The goal of training the network is to maximize the scale-invariant signal-to-noise ratio (SI-SNR), which is defined as:

$$SI - SNR = 10 \log_{10} \frac{\|s_{\text{target}}\|^2}{\|e_{\text{noise}}\|^2}$$

Among them, $\hat{s} \in \mathbb{R}^{1 \times T}$ and $s \in \mathbb{R}^{1 \times T}$ are the estimated source signal and the original source signal, respectively, and $\|s\| = \langle s, s \rangle$ represents the signal power. Normalize $\hat{s}$ and $s$ to zero mean before calculation to ensure that the scale remains unchanged, and avoid reducing the signal-to-noise ratio due to the difference between the amplitude of the estimated source signal and the original source signal. During training, the uPIT method is used to solve the source replacement problem, and the estimated signal is one-to-one corresponding to the original signal.

4.4. Evaluation index
In order to evaluate the effect of blind source separation, this paper selects three evaluation indicators for evaluation, namely waveform similarity coefficient, error rate, and scale-invariant signal-to-noise ratio (SI-SNR). Substitute the initial source signal and the estimated source signal into the following equation to perform calculations to obtain the calculation result of the evaluation index, thereby judging the effect of blind source separation. In order to compare the superiority of the ConvTasNet method, this paper selects the traditional blind source separation algorithm FastICA to blind source separation of the simulated signal and the actual transformer mixed signal, and compares and analyzes it[3].
4.4.1. Waveform similarity coefficient.

\[ \lambda_{ij} = \lambda(y_i, s_j) = \frac{\sum_{k=1}^{K} y_i(k)s_j(k)}{\sqrt{\sum_{k=1}^{K} y_i^2(k)} \sqrt{\sum_{k=1}^{K} s_j^2(k)}} \]  

(7)

In the formula, \( \lambda \) represents the similarity relationship between signals \( y_i \) and \( s_j \), and the value range is \( 0 \leq \lambda \leq 1 \). The larger the value of \( \lambda \), the higher the similarity of the waveform. When \( \lambda = 1 \), the two waveforms are completely equal.

4.4.2. Scale-invariant signal-to-noise ratio (SI-SNR).

\[ SI-SNR = 10 \log_{10} \frac{\| s_{\text{true}} \|^2}{\| e_{\text{noise}} \|^2} \]  

(8)

4.5. Experimental results and analysis

Select a sample in the test set, and the separation result is shown in the figure below. In order to better display the time-domain waveform, only 1500 points in the sample are drawn in the time-domain diagram.

Fig. 3 is a mixed signal of two simulated source signals, which are a time-domain diagram and a frequency-domain diagram respectively. It can be seen from the figure that the frequency of the simulated mixed signal is within 1kHz, which is a frequency multiplication component of 50Hz.

In Fig. 4, the first line is the time-domain diagram and frequency-domain diagram of the two source signals, and the second line is the time-domain diagram and frequency-domain diagram of the estimated signal. It can be seen from the figure that the estimated signal and the source signal have similar waveforms and the corresponding frequency bands are basically the same, indicating that the algorithm has a better effect of estimating the source signal. According to the calculation, the waveform similarity coefficient is 0.921 and the SI-SNR=8.544.

In order to compare the separation effect of this method, this article uses the traditional FastICA-based blind source separation algorithm to compare with the method in this article. ConvTasNet is a single-channel blind source separation algorithm, and FastICA is a positive definite blind source separation algorithm. Separating two source signals requires two mixed signals. The result of FastICA blind source separation is shown in Figure 5 and Figure 6.

![Figure 3. Time domain diagram and frequency domain diagram of mixed signal.](image-url)
Figure 4. Time domain diagram and frequency domain diagram of source signal and estimated signal.

Figure 5 is a signal time domain diagram, and Figure 6 is a signal frequency domain diagram. In the two figures, the first row is the mixed source signal obtained by randomly weighting the two source signals, the second row is the two source signals, and the third row is the two estimated signals obtained by the FastICA algorithm. It can be seen from the figure that the separation effect of source signal 1 is good, 100Hz, 300Hz, 400Hz components can be separated well, and the separation effect of source signal 2 is poor. The estimated signal 300Hz, 400Hz and the source signal are far apart After calculation, the waveform similarity coefficient is 0.723 and SI-SNR=5.635. Compared with the two algorithms, the ConvTasNet algorithm has a better separation effect on two source signals with strong correlation. It can better separate the same frequency bands in the two source signals, and the waveform similarity coefficient and SI-SNR indicators are better. Excellent.

Figure 5. Time domain diagram of signal.
5. Conclusions
In this paper, the blind source separation algorithm based on ConvTasNet is applied to separate the winding and iron core vibration signals from the mixed signal on the surface of the transformer tank. Based on this, the simulation experiment is carried out to verify. Through experiments, it can be concluded that based on this algorithm, the relatively strong winding and core signals can be better separated, and the waveform similarity coefficient and SI-SNR are evaluated, and the signal separation effect is better than the traditional FastICA blind source separation algorithm. It provides an idea for monitoring the operating status of transformer windings and cores based on vibration signals, but further work is needed for experimental verification and field applications based on actual vibration signals.

Acknowledgments
This work is supported by the Science and Technology Project of State Grid Corporation of China under Grant No. 520201190090.

References
[1] BAGHERIM, NADERIM. (1979) On-line transformer winding deformation diagnosis: a Profound insight to methods.C. Proceedings of 26th International Power System Conference, Tehran, Iran.pp. 1-14.
[2] LORDT, HODGEG USA. (2002) Interfacing on-line monitoring technologies to power transformers. http://www.lordconsulting.com/images/stories/pdf/Transformer-Monitoring-jul10.pdf.
[3] KOCI SKIE, SZYMOND, EDWARDO.(2012) Evaluation of blind source separation for different algorithms based on second order statistics and different spatial configurations of directional microphones. J. Applied Acoustics, 2012, 73(2): 109-116
[4] Isik, Y., et al. (2016) Single-Channel Multi-Speaker Separation Using Deep Clustering. 17TH ANNUAL CONFERENCE OF THE INTERNATIONAL SPEECH COMMUNICATION ASSOCIATION (INTERSPEECH 2016). Location San Francisco.pp.545-549.
[5] Luo, Y., and N. Mesgarani. (2019) Conv-TasNet: Surpassing Ideal Time–Frequency Magnitude Masking for Speech Separation. IEEE/ACM Transactions on Audio, Speech, and Language Processing PP.99(2019):1-1.