Power Transformer Voiceprint Operation State Monitoring Considering Sample Unbalance

ShouLong Chen¹, Ping He¹, HongHua Xu¹*, LaiBin Yin¹, LingYan Wang¹ and Lei Zhu¹

¹State Grid Jiangsu Electric Power Co., Ltd. Nanjing Power Supply Branch, Nanjing, Jiangsu, 210019, China

*Corresponding author’s e-mail:1191341704@qq.com

Abstract. The voiceprint characteristics of transformers are closely related to the operating conditions, but there is currently a lack of effective research on the voiceprint characteristics of transformers during operation. First of all, this article collects three operating conditions of load, light load, and no load on the basis of the transformer voiceprint signal acquisition platform. Secondly, in view of the characteristics of the transformer's voiceprint, the 50Hz frequency multiplier component amplitude is extracted to form a feature vector, which solves the problem of low utilization rate of common feature extraction information. Finally, in view of the problem of transformer voiceprint failure and sample imbalance caused by fewer abnormal samples, a pattern recognition based on the RUSBoost algorithm is proposed. The algorithm has good recognition accuracy and applicability for transformer voiceprint samples with imbalance problems. The research results provide effective support for the monitoring and identification of the mechanical condition of transformers with sample unbalanced voiceprints, and the analysis of the operating conditions can effectively eliminate the errors that may be caused by their own different operating conditions.

1. Introduction

Transformer is one of the core equipment of power system. Its safe and stable operation play an important role in the safety and reliability of power system[1-2]. For a long time, people have always regarded the sound made by the transformer as noise while ignoring the large amount of equipment information contained in it. Transformer state monitoring and fault diagnosis based on sound signal, because it has no contact with the measured object, and it is easy to realize live monitoring and diagnosis, thus showing broad application prospects[3-4].

In recent years, with the development of artificial intelligence technology, due to its high efficiency and high accuracy in prediction, classification, etc., this provides a new way of thinking for transformer fault diagnosis and condition monitoring. Aiming at the poor positioning accuracy of conventional beamforming algorithms, Bao Hailong, Shao Yuying and others proposed a dry-type transformer abnormal noise fault study based on deconvolution beamforming algorithm[5]. Liu Yunpeng, Luo Shihao, etc. proposed a method for identifying loose voiceprints of transformer cores based on Mel time-frequency spectrum-convolutional neural network[6]. Ma Wenjia, Wang Fenghua and others proposed a transformer winding state detection method based on sparse adaptive S-transform for the acoustic signal when the transformer is subjected to a short-circuit impact[7]. Large power transformers have high reliability and stable working conditions. It is often difficult to obtain fault or abnormal data. Traditional research methods can perform effective feature extraction and pattern recognition on transformer
voiceprint signals, but they have poor applicability to unbalanced data. Data sample requirements are higher and other issues.

In summary, this paper proposes a power transformer voiceprint operation status monitoring considering sample imbalance. First of all, on the basis of the transformer voiceprint data acquisition platform, the voiceprint data of three operating states of load, light load and no load are collected. Secondly, according to the characteristics of the transformer's voiceprint, the 50Hz frequency multiplier component amplitude is extracted as the characteristic component, which solves the problem of low information utilization in traditional methods. Finally, the RUSBoost algorithm is used to identify the operating conditions of the unbalanced voiceprint signal sample data.

2. Analysis of the mechanism and conduction process of transformer voiceprint
Transformer voiceprint can generally be divided into iron core voiceprint and winding voiceprint. The sound ripple of the transformer is produced by the vibration of the iron core and the winding. The sound wave propagates through a liquid or solid path, and the sound wave that finally propagates to the outside of the transformer is a combination of multiple sound sources. The sound wave signal has very little attenuation in the transformer oil and air transmission, and the sound wave signal will not have frequency deviation when passing through the tank wall. Therefore, the voiceprint characteristics are basically the same as the vibration characteristics. This article analyzes the voiceprint of the transformer by analyzing the vibration principle.

2.1. Analysis of iron core acoustic texture theory
The iron core vibration sound pattern is mainly caused by the magnetostrictive effect of the silicon steel sheet. Therefore, the acceleration caused by magnetostriction can be obtained as:

\[
a = \frac{2\varepsilon LU_0^2}{(N_1AB_s)^2} \cos 2\omega t
\]

(1)

It can be seen that the vibration frequency of the silicon steel sheet in the alternating magnetic field is twice the voltage frequency. For a 50Hz power grid, 100Hz is the fundamental frequency of the transformer core vibration. When the tightening force is not uniform, the silicon steel sheet will produce periodic reciprocating motion, resulting in an even frequency component of 50 Hz.

2.2. Analysis of winding acoustic texture theory
The winding is the basic structure of the transformer, and the coil is the basic unit of the winding, and there are many winding methods. The previous mass-spring-damping models are mostly aimed at the pie structure, which is not universal. The wire between the spacer blocks is used as the basic physical unit, which is called the simplified coil basic unit. This physical model with mechanical characteristics and electromagnetic field characteristics is called the winding two-body model.

\[
x = \frac{\mu l^2}{2\pi r_0 K} + \frac{\mu l^2}{2\pi r_0 (K - 4\omega^2 M)} \cos(2\omega t)
\]

(2)

The first term of the equation is the constant component produced by the constant force, and the frequency of the winding vibration in the second term is 2ω. For a 50Hz power system, the vibration frequency is 100Hz.

2.3. Pattern recognition based on RUSBoost algorithm
The distribution of transformer samples is often unbalanced, with more data in the normal operating state and less data in the non-normal operating state. This makes the model prediction results tend to be biased toward the normal state, and the effect of identifying more meaningful abnormal samples is not good. Aiming at the serious imbalance problem of sample data, this paper proposes to build a transformer voiceprint recognition model based on the RUSBoost algorithm, improves the balance of the data sample through undersampling, and combines the lifting method to upgrade multiple simple
base learners to strong learners. Compare the training error of the current base learner to adjust the distribution weight of the training sample, and increase the penalty factor to increase the attention in the subsequent training process, and then use the adjusted sample to train the next base learner, and iterate repeatedly. The flow chart of transformer voiceprint operation monitoring based on 50FMWTE-RUSBoost is shown in the figure below. The specific steps are as follows:

1) 50Hz double frequency wavelet time-frequency entropy calculation and feature extraction. The wavelet coefficients of each doubling frequency of 50 Hz are extracted to calculate the entropy value, and the maximum value of the amplitude is used to calculate the weight of each component.

2) Build the RUSBoost algorithm recognition model. Collect and construct training set data samples, deal with unbalanced data through under-sampling, and build a pattern recognition model combined with the lifting method.

3) Transformer operating status monitoring based on voiceprint. Calculate its 50Hz doubled wavelet time-frequency entropy and bring it into the trained RUSBoost algorithm recognition model to obtain its specific operating status information.

![Algorithm flowchart](image)

**Figure 1. Algorithm flowchart**

### 3. Platform construction and data analysis

#### 3.1. Transformer Voiceprint Data Acquisition Platform

This paper builds a 220kV autotransformer voiceprint data acquisition platform, as shown in Figure 1. The data acquisition platform is mainly composed of 220kV autotransformer, computer, DHDAS dynamic signal acquisition instrument, signal transmission line, preamplifier HS14618 and capacitive acoustic sensor HS14018 Wait. The measuring points are arranged as shown in Figure 2:

![Sound measuring point](image)

**Figure 2. Substation data collection**

According to the international measurement standard IEC60651. The acoustic signal measurement should cover the audible sound range of 20Hz-20kHz, so the equipment uses a sampling frequency of 50kHz. The use of signal transmission lines that are resistant to strong magnetic field interference can effectively reduce external electromagnetic field interference. The capacitive acoustic sensor is 2m away from the outer wall of the transformer tank and 1.5m away from the ground.
3.2. Voiceprint signal time-frequency domain and measurement point location analysis

The frequency spectrum and the time frequency spectrum are important characteristic frequency spectrums of acoustic signal processing. As shown in the Figure 3, the frequency components of the voiceprint signal during the normal operation of the transformer appear as 50Hz, 100Hz, 150Hz and other 50Hz double frequency components, of which 100Hz, 200Hz and 300Hz frequency components are relatively large. This paper compares the data of different measuring points. The position of the measuring point is different, and the amplitude of each frequency component of the sound signal is different, but they all show similar trends. The sound signal is generally within the range of 900 Hz, and the sound signal The frequency is dominated by even multiples of 50Hz, and 50Hz and its odd multiples account for less.

![Wavelet time-frequency diagram](image)

![Spectrogram](image)

Figure 3. Spectrogram and time spectrum

The main frequency of the transformer's voiceprint is 100Hz under load, light load and no-load conditions. In the load state, the current flowing through the winding is relatively large, and the voiceprint of the winding dominates at this time. In the light load state, the current decreases, and the voiceprint is determined by the voiceprint of the winding and the voiceprint of the iron core. In the no-load state, no current flows in the windings, and the voiceprint is mainly determined by the voiceprint of the iron core. As shown in Figure 4.

![Figure 4. Spectrogram of different working conditions](image)

4. Condition monitoring and pattern recognition

4.1. Transformer Voiceprint Operation State Monitoring and Pattern Recognition

Aiming at the characteristics of the voiceprint signal of the transformer, a voiceprint feature extraction method based on 50Hz doubling wavelet time-frequency entropy is proposed, which not only ensures the extraction of key information, but also prevents the amount of data from being too large. Improve the random forest algorithm, optimize and adjust its parameters, and improve the recognition rate.

In order to ensure the effectiveness of model recognition, during model training, the sample set needs to be divided into training set and test set. The sample division is shown in Table 1 below. In a total of 542 sets of samples, there are 257 sets of normal load samples, 240 sets of light load samples and 45 groups of unloaded samples. The maximum imbalance rate among samples of each category is 5.71. Set
labels for different working conditions in the table, and finally all labeled samples are randomly input into the model for training.

The model is trained and tested ten times and passed through the grid. The search method optimizes the hyperparameters. Since the average accuracy rate is not comprehensive enough to evaluate the performance of unbalanced samples, the optimization goal of the grid search is to find the optimal AUC value. The final setting adjusts the maximum number of splits to 37, the number of base learners to 45, and the learning rate to 0.12. The evaluation result can be represented by the confusion matrix as shown in the figure 5.

![Confusion Matrix](image)

In the confusion matrix shown in Figure 5, the abscissa is the predicted result of the model, the ordinate is the true result of the model, and the rightmost column is the correct rate and error rate of model recall. It can be seen from the figure that the recall rate of all samples is above 94%, which indicates that the model is very accurate in identifying the normal samples with a large number of samples and the samples with a small number of samples. The overall accuracy reaches 98.9%, and the AUC value is 0.98, indicating that the model has good recognition accuracy and applicability for transformer voiceprint samples with unbalance problems. In addition, the common Decision Tree (DT), Random Forest (RF), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM) algorithms are compared with the algorithms proposed in this article. Lattice search finds the best hyperparameters to make its AUC value the best, and the training model results are shown in the figure 6.

![Algorithm Accuracy](image)
It can be seen from the analysis that the recall rate of no-load samples of RF is 92.1%, which is 2.5% lower than that of the RUSBoost model. The recall rate of light-load samples of DT is 82.4%, which is 14.8% lower than that of the RUSBoost model. The recall rate of no-load samples of KNN is 86.6%. Compared with the RUSBoost model, the recall rate of no-load samples of SVM is 52.9%, which is 41.7% lower than that of the RUSBoost model. It can be seen that the RUSBoost model has a better classification accuracy rate for unbalanced transformer sample data. It plays an important role in solving transformer faults or missing abnormal sample data.

5. Conclusions
This paper takes the transformer voiceprint signal as the research object, collects data through the transformer voiceprint data acquisition platform, studies the different working conditions of the transformer during operation which is a transformer based on voiceprint. Provide basis for condition monitoring and fault diagnosis. The main conclusions in the article are as follows:

1) Through theoretical research and experimental data analysis, the transformer's voiceprint signal is mainly a 50Hz doubling frequency component. Aiming at this feature, a 50Hz doubling wavelet time-frequency entropy is proposed, which improves the utilization and abundance of information.

2) Aiming at the problems of unbalanced transformer voiceprint samples, faults and fewer abnormal samples, a RUSBoost-based transformer voiceprint operating mode recognition is proposed, with an overall accuracy of 98.9%, and the recall rate of no-load samples with a small sample size 94.6%, a maximum increase of 41.7% compared with other traditional classification algorithms. The results show that the RUSBoost algorithm has better accuracy for unbalanced sample data than common classical algorithms.

Acknowledgments
Key Technology Project of State Grid Jiangsu Electric Power Co., Ltd. (J2021053)

References
[1] Liang Deliang, Liu Yibin, Kou Peng, et al. Analysis on the development trend of intelligent distribution transformers[J]. Automation of Electric Power Systems, 2020, 44(07): 1-14.
[2] Qi Bo, Zhang Peng, Zhang Shuqi, et al. Application status and development prospects of digital twin technology in condition assessment of power transmission and transformation equipment[J]. High Voltage Technology, 2021, 47(05): 1522-1538.
[3] Li Peng, Bi Jiangang, Yu Hao, et al. Intelligent sensing and state perception technology and application of substation equipment [J]. High Voltage Technology, 2020, 46(09): 3097-3113.
[4] Pu Tianjiao, Qiao Ji, Han Xiaojie, et al. Research and application of artificial intelligence technology in the operation and maintenance of power equipment [J]. High Voltage Technology, 2020, 46(02): 369-383.
[5] Wang Fenghua, Wang Shaojing, Chen Song, et al. Transformer voiceprint recognition model based on improved MFCC and VQ[J]. Proceedings of the Chinese Society of Electrical Engineering, 2017, 37(05): 1535-1543.
[6] Geng Qishen, Wang Fenghua, Jin Xiaojie. Dry-type transformer mechanical fault sound diagnosis based on Gammatone filter cepstrum coefficients and whale algorithm optimization random forest [J]. Electric Power Automation Equipment, 2020, 40(08): 191-196+224+197-199.
[7] Liu Yusheng, Wang Bowen, Yue Haotian, Gao Fei, Han Shuai, Luo Shihao, Zhang Chenchen. Transformer bias voiceprint recognition based on 50Hz double frequency cepstrum coefficient and gated loop unit[J]. Proceedings of the Chinese Society of Electrical Engineering, 2020, 40 (14):4681-4694+4746.