Real time outlier monitoring for power transformer fault diagnosis based on isolated forest

Li Shen¹, Hongjun Du¹, Shuji Liu¹, Shuo Chen¹, Lin Qiao¹, Sai Liu², Jiahua Liu², Kexin Li³ and Jing Li ³,*

¹State Grid Liaoning Electric Power Supply CO, LTD. Shenyang, 110004 Liaoning, China
²NARI Group Corporation/State Grid Electric Power Research Institute, Nanjing, 211000 Jiangsu, China
³College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, 211106 Jiangsu, China

*E-mail: lijing@nuaa.edu.cn

Abstract. In order to improve the accuracy and efficiency of transformer fault detection, this paper uses the isolated forest algorithm combined with the historical transformer characteristic gas data to establish the characteristic gas outlier recognition model, and then uses the uncoded ratio method to establish the abnormal event strategy and abnormal event library for the transformer historical fault information. Finally, the state of the transformer is diagnosed based on the established outlier detection model and the exception event library. The experimental results show that the proposed method has a great improvement in the detection efficiency and stability of outliers, and is more accurate in transformer fault diagnosis combine with abnormal event database.

1. Introduction

With the continuous development of artificial intelligence, the scale of the power grid in the power industry is also constantly increasing, and the requirements for power quality of users are also constantly improving. The safe and stable operation of equipment in the power grid is a prerequisite for meeting the needs of users [1]. However, in the actual operation process of the power equipment, certain faults will always occur due to the equipment itself or external factors. If the equipment fails, it will bring irreparable damage to the grid operation.

Transformers are the key equipment for power system energy conversion and transmission. They often work in various stress environments, such as electricity, heat, high voltage, high current, moisture, and oxygen. The internal main insulation properties will suffer irreversible aging and eventually fail. Failure to find and take appropriate measures in time will result in failure of the transformer and unnecessary economic losses. Therefore, timely grasping the operating state of the transformer equipment has special significance for the safe operation of the power supply system [2].

As a method to identify the internal discharge and overheat abnormality of oil-immersed transformers, the dissolved gas analysis [3] has become a means of transformer anomaly detection. Since there are many hydrocarbons with different molecular weights in the transformer oil, the molecule contains chemical groups such as CH₃, CH₂, CH, and C-C bonds and C-H bonds. Under the
action of electricity or heat, some C-C bonds and C-H bonds are broken to form unstable hydrogen atoms and hydrocarbon radicals, which are re-synthesized into hydrogen and low molecular hydrocarbon gases. Due to the different degrees of oil temperature and discharge, the gases produced are also different. Therefore, the dissolved gas method in the oil can detect and diagnose the abnormality of the transformer according to the dissolved gas composition and the content of the transformer oil [4].

2. Related work
As a common method for detecting the abnormal state of the transformer, the dissolved gas method in the oil is used as the statistical threshold by the dissolved gas content in the oil. The threshold value is derived from the statistics of a large number of transformer historical characteristic gas detection data, and is judged to be abnormal when the monitored value exceeds the specified threshold data. The key to this method is that for the threshold setting, when the historical data of the transformer is large, the method can be well recognized, but if the historical data is lacking, the detection effect of the method will be greatly reduced. In response to the problems of the statistical threshold method, many scholars began to study other methods to detect the abnormal state of the transformer. Such as based on characteristic gases [5-7], based on IEC three ratios [8-10], based on the non-coding ratio method [11].

Based on the characteristic gas method, fault diagnosis is performed based on the difference in gas production rate, gas type, and gas concentration [6] of the transformer. Although this method can intuitively and conveniently detect the abnormal state of the transformer, there is no quantitative concept, and information such as temperature change law is needed to accurately locate the fault type. The three-ratio method selects two gases with similar solubility coefficient and diffusion coefficient from C$_3$H$_2$, C$_2$H$_4$, CH$_4$, H$_2$ and C$_2$H$_6$ gases according to the dependence of the relative characteristics of different characteristic gases on the transformer temperature during transformer failure, forming three contrast values [8]. Transformer fault diagnosis is performed according to different coding combinations by different coding representations. Although this method has a quantitative concept, the calculation is relatively simple. However, certain conditions are required for use, and there are cases where there is no corresponding failure in the ratio. On this basis, a codeless rate method [11] is proposed. This method solves the problem of fault coding of three-ratio method to some extent, and the problem that some three-valued method fault cannot be diagnosed. However, this method also diagnoses transformer faults based on the ratio between the corresponding characteristic gases, and does not fundamentally solve the problem of lack of ratio coding between the characteristic gases. Later, based on support vector machine [12], extreme learning machine [13], deep learning [14] and other machine learning methods are also used for transformer fault diagnosis. Although the machine learning-based approach improves the efficiency and accuracy of transformer fault diagnosis to a certain extent, these methods rely on transformer fault data. When a new fault occurs, only the existing empirical data does not identify the fault well.

3. The proposed solution
In this paper, firstly, based on the historical transformer characteristic gas data, the isolation forest detection algorithm is used to establish the outlier detection model, and then the current transformer characteristic gas real-time data is detected, and the detected abnormal value is combined with the transformer abnormal event library to judge the current transformer state. The overall steps for transformer outlier monitoring and abnormal state diagnosis based on isolated forests are shown as figure 1:
3.1. Outlier detection model based on isolated forest

The key to the Isolation Forest algorithm is the construction of the isolation tree. First, the original data is used to establish the isolation tree, and then the isolation tree is combined into an isolated forest for the identification and detection of outliers. In the isolated forest, in order to detect outliers, it is necessary to define the calculation method for the isolation tree and path length.

Isolation Tree: An isolation tree can be thought of as a binary search tree. If \( T \) is a node of the isolation tree, then \( T \) is either a leaf node or a child node \((T_L, T_R)\) with a test instance. The test at the node \( T \) consists of the attribute \( p \) and the segmentation value \( q \). The segmentation attribute and the segmentation value are used to segment different data, the data record smaller than the segmentation value is divided into the left child node, and the data record larger than the segmentation value is divided into the right child node. Repeat the above process until there is only one data in the child node or the maximum height of the tree has been reached. It is generally considered that the data corresponding to the shorter path is recorded as an abnormal value, and the record corresponding to the longer path is a normal value.

Path length: The path length \( h(x) \) of the data record \( x \) refers to the number of edges encountered during the traversal process from the root node to the traversal of the isolation tree until the leaf node is encountered.

Outlier score: Since the constructed isolation tree and the binary search tree have the same structure, here we use the average length that was not successfully searched in the corresponding binary search tree as the normalization factor. From the binary search tree, for a given sample size of sample size \( \varphi \), the average length of the corresponding binary search tree is:

\[
c(\varphi) = \begin{cases} 
2H(\varphi - 1) - 2(\varphi - 1) / \varphi & \text{if } \varphi > 2, \\
1 & \varphi = 2, \\
0 & \text{otherwise.}
\end{cases}
\]  

(1)

where, \( h(\varphi) \) is the harmonic function, which can be calculated by \( \ln(\varphi) + 0.5772156649 \) (Euler constant). After the normalization factor is obtained, we normalize the path length, and then calculate the anomaly of data \( x \) is worthy of classification. The anomaly is worthy of classification as follows:

\[
s(x, \varphi) = \frac{E(h(x))}{c(\varphi)}
\]  

(2)

\[
E(h(x)) = \frac{1}{n} \sum_{i=1}^{n} h_i(x)
\]  

(3)

where \( h(\varphi) \) represents the path length of the data \( x \) on the \( i \)-th tree, and \( E(h(x)) \) represents the average of the path lengths of the data \( x \) in different isolation trees.

The Isolation Forest algorithm can be divided into two steps when performing outlier detection: the first step, the training process. The original data set is randomly sampled to obtain multiple sub-data
sets, the sub-data set is used to construct the isolation tree, and then the constructed isolation tree is used to form the isolation forest. The second step, the prediction process, can also be called the outlier scoring process. The data of the detected isolated forest is used to calculate the outlier score, and the outlier score is compared with the set threshold to judge the abnormal data. For the calculation method of the abnormal value score established above, it can be known that when all the abnormal values of the data to be detected are scored. When both are close to 0.5, it means that the data set has no obvious outliers; when \(E(h(x))\) is closer to 0, the outlier value \(s\) is closer to 1, which means that the data is basically regarded as abnormal data; when \(E(h(x))\) is closer to the sample size \(\varphi\), the outlier score is closer to 0. If the outlier score is much smaller than 0.5, the data is considered to be normal data.

3.2. Fault diagnosis model

The outlier identification model based on the isolated forest can detect the abnormal value in the characteristic gas data of the transformer, but cannot determine whether the state of the transformer corresponding to the abnormal value is normal. Therefore, this paper proposes a real-time diagnosis of transformer faults based on the model of anomaly strategy and anomalous event library.

1) Abnormal state strategy

An abnormal state identification library is established according to the content of the characteristic gas and the proportion of the characteristic gas, and the characteristic gas abnormal data and the abnormal state of the transformer are matched to achieve the purpose of identifying the abnormal state of the transformer. According to the change of the characteristic gas value, an abnormal state strategy is formulated, and the judgment criteria are as follows:

(a) When the values of all the characteristic gases are simultaneously decreased or increased, it is judged that there is no abnormality;

(b) When the value of some characteristic gases increases and the proportion increases or decreases significantly, it is judged to be abnormal;

(c) When the content of one or several gases is increased and the combined characteristic gas combination can be matched with the abnormal state in the abnormal state library, it is determined to be an abnormal state;

(d) When one or several characteristic gases increase and do not match all events in the abnormal state library, it is judged to be an abnormal state and relevant experts are required to perform auxiliary judgment.

2) Abnormal state event library

The abnormal state event library is a combination of typical characteristic gases corresponding to various abnormalities of the transformer. The abnormal event library can be used to accurately analyze the abnormal state of the transformer. The main body of the abnormal state event library is the characteristic gas, and the abnormal event library can be continuously enriched according to the actual situation. The exception status event library is shown in Table 1:

| Abnormal state event | Characteristic gas | \(C_2H_2/C_2H_4\) | \(C_2H_4/C_2H_6\) | \(CH_4/H_2\) | Abnormal state event |
|----------------------|--------------------|-------------------|------------------|--------------|---------------------|
| H2                   | CH4                | 0.1               | <1               | T1           | Irrelevant          |
| O                    | 2                  | 1                 | —                | 1            | T2                  | Relevance<1&<3 |
| OP                   | 2                  | 1                 | —                | 1            | T3                  | Relevance<1&<3 |
| PD&OP                | 1                  | 1                 | 2                | 0.1          | T4                  | Relevance<1&<3 |
| SDO                  | 1                  | 1                 | 2                | 0.1          | T5                  | Relevance<1&<3 |
| OA                   | 1                  | 2                 | 2                | 0.1          | T6                  | Relevance<1&<3 |
| AOP                  | 1                  | 2                 | 2                | 0.1          | T7                  | Relevance<1&<3 |

Table 1. Abnormal state event library.
Note: 1 means the main gas, 2 means the secondary gas, - means irrelevant, O means oil overheating, OP means oil and insulation paper overheated, PD&OP means partial discharge in oil and paper insulation, SDO means spark discharge in oil, OA means oil arc, AOP means arc in oil insulating paper, T1 means thermal fault of T<300 °C, T2 means thermal fault of 300°C<T<700°C, T3 means thermal fault of T >700°C, D1 means high energy discharge, D2 means low energy discharge, T means thermal fault.

4. Experiment
This paper takes the characteristic gas history test data of No. 3 oil-immersed transformer of 220KV substation in a certain area of Nanjing as an example. Firstly, the time domain analysis of the characteristic gas data of the above transformer is carried out to obtain whether the characteristic gas is abnormal or not.

As shown in figure 2, it can be seen from the results of the above time domain diagram that there is a significant step in the characteristic gas data of the transformer, indicating that there is an abnormal value in the characteristic gas, and the state of the transformer may be abnormal. The transformer characteristic gas data is modeled by an outlier detection algorithm based on the isolated forest to identify outliers in the characteristic gas. According to the historical characteristic gas data of the transformer, the isolation forest is established. The detection effect of the algorithm on the historical data is shown in the figure 3.

According to the outlier detection model established by the isolated forest algorithm on the characteristic gas of the transformer, the characteristic gas data of other transformers are detected, and the abnormal data of the characteristic gas of the detected transformer is used to diagnose the fault of the transformer. Using the same transformer characteristic gas data, the method proposed in this paper is compared with other transformer abnormality diagnosis methods. The results are shown in table 2:
Table 2. Detection results and comparison with other methods.

| Time Point | Guide method | Method of this paper | Transformer real state |
|------------|--------------|----------------------|------------------------|
| 1          | abnormal     | Discharges of High Energy | Discharges of High Energy |
| 2          | abnormal     | Discharges of Low Energy | Discharges of Low Energy |
| 3          | abnormal     | Thermal fault         | Thermal fault           |
| 4          | normal       | normal                | normal                 |
| 5          | abnormal     | Thermal fault T>700℃ | Thermal fault T>700℃   |
| 6          | normal       | Partial Discharge     | Discharges of Low Energy |
| 7          | normal       | normal                | normal                 |
| 8          | abnormal     | Thermal fault 300℃<T<700℃ | Thermal fault 300℃<T<700℃ |
| 9          | abnormal     | Discharges of High Energy | Discharges of High Energy |
| 10         | abnormal     | Thermal fault T<300℃ | Thermal fault T<300℃   |

Through the above test results, it can be found that compared with other transformer abnormal state diagnosis, the isolated forest-based transformer outlier detection method and fault diagnosis model proposed in this paper can not only detect the transformer with or without fault, but also accurately detect the transformer fault type.

5. Conclusion

Based on the isolated forest algorithm, a transformer characteristic gas outlier recognition model and transformer fault detection method based on anomaly strategy and abnormal event library are proposed. Firstly, the characteristic gas data of the transformer is modeled by isolating the forest, and the abnormal value identification of the characteristic gas of the transformer is completed. Then, the abnormal value identified is classified according to the abnormality strategy. Finally, the classification result is performed according to the abnormal event library to diagnose the type of fault in the transformer. When the fault of the transformer is diagnosed, the method proposed in this paper only diagnoses the data with abnormality. This method can reduce the computational cost in the diagnosis process. Combined with the exception event library, the specific fault type of the transformer can be detected more accurately.

Acknowledgments

The research was supported by the State Grid Liaoning Electric Power Supply CO, LTD. and we are grateful for the financial support for the “Key Technology and Application Research of the Self-Service Grid Big Data Governance (SGLNXT00YJJS1800110)”.

References

[1] F. Wan, W. Chen, X. Peng, et al. Study on the gas pressure characteristics of photoacoustic spectroscopy detection for dissolved gases in transformer oil. International Conference on High Voltage Engineering and Application, 286-89(2012)
[2] X. Zhao, Y. Li. An On-Line Monitoring System for Gases Dissolved in Transformer Oil Using Wireless Data Acquisition. Power and Energy Engineering Conference, 1-4(2009)
[3] X. Liu. Application of Dissolved Gas Online Monitoring System in Transformer Oil in Gezhouba Power Station. Hydropower & Pumped Storage (2017)
[4] J. Dai, H. Song, G. Sheng, et al. Dissolved gas analysis of insulating oil for power transformer fault diagnosis with deep belief network. IEEE Transactions on Dielectrics & Electrical Insulation, 24(5), 2828-35(2017)
[5] P. Wang, W. Chen, F. Wan, et al. Simultaneously analyze fault characteristic gases extracted from transformer oil by Raman spectroscopy. IEEE International Conference on High Voltage Engineering and Application, 1-4(2016)
[6] W. Chen, L. Zhao, S. Peng, et al. Analysis of Dissolved Gas in Transformer Oil Based on Laser Raman Spectroscopy. Proceedings of the Csee, 34(15), 2485-92(2014)
[7] D. Yang, X. Chen, Z. Ma. Fault detection of transformers based on Raman spectra of the dissolved gas in transformer oil. Power Electronics and Motion Control Conference, 2431-36(2016)
[8] X. T. Zhao, Q. Shen, X. U. Jia-Xiang, et al. Fault Diagnosis for Power Transformer Based on IEC Three-ratio and Extension Association with Combined Weights. Power System & Clean Energy, (2013).
[9] P. U. Lijuan, N. Liu, H. Liu, et al. Statistical Analysis of IEC Three-ratio Method Based on Fault Data of Oil-immersed Transformers. Shanxi Electric Power( 2015)
[10] L. F. Lin, Z. Q. Wen, S. C. Deng, et al. Research on Fault Diagnosis for Power Transformer Based on Fuzzy Three-Ratio Method and Case-Based Reasoning. Transformer (2015)
[11] H. Xue, H. P. Zhang. Fuzzy No Coding Ratio in Transformer Oil Gas Detection in the Application. Science Technology & Engineering (2012).
[12] Y. Zhang, B. Zhang, Y. Yuan , et al. Transformer fault diagnosis based on support vector machine. Telecom Power Technology, 6, 405-08(2016)
[13] L. L. Wang, F. Pei, Y. L. Zhu. Transformer Fault Diagnosis Based on Online Sequential Extreme Learning Machine. Applied Mechanics & Materials, 721, 360-65(2015)
[14] T. Yang, J. Huang, X. U. Kui, et al. Diagnosis method of power transformer fault based on deep learning. Power Systems & Big Data, (2018)