Research on Transformer State Evaluation Method Based on Fault Feature Extraction

Shuang Liua, Rui Han, Yongtao Jin, Wenhao Wang, Haofan Lin and Zhi Yang

State Grid Zhejiang Electric Power Research Institute, Hangzhou 310014, China

*Corresponding author’s e-mail: 18729562087@163.com

Abstract. In order to realize the dynamic evaluation of transformer status and gradually improve the level of equipment fault diagnosis, a transformer condition assessment method based on fault feature extraction was proposed in this paper. Based on the data of the key state quantity of dissolved gas in transformer oil, this method extracted the transformer fault characteristics through various mathematical methods such as hypothesis testing and logistic language regression, and selected new indexes that could better reflect the operating state of transformer, so as to judge the operating state of transformer. In this paper, a large number of indicators were screened, and the regression method was used to find the indicators to judge the fault, without other external conditions. The analysis results showed that the average prediction accuracy of the proposed method was more than 95% under multiple cross validation, which had high engineering application value.

1. Introduction

The power transformer is the core equipment of the power grid. It plays an immeasurable role of continuity and transformation in the power grid. Its health state directly affects the safe and stable operation of the system [1].

Transformers mainly use insulating oil (liquid), insulating oil paper, laminated cardboard (solid) and other insulating materials as the insulation system. These insulating materials in the operation of the transformer by temperature, electric field and catalyst, will crack to produce some specific gas called characteristic gas. The characteristic gas dissolved in transformer oil can reflect the thermal decomposition nature of oil and paper insulation caused by fault point, which has become one of the most effective means to diagnose transformer fault at home and abroad [2]. Among them, the most commonly used method in engineering applications is the three-ratio method recommended by IEC. The characteristic gas three-ratio method is to calculate the C2H2/C2H4, CH4/H2 and C2H4/C2H6 three groups of characteristic gas concentration ratios when the characteristic gas concentration exceeds the attention value to obtain the corresponding code, and then judge the transformer fault by the code type. This method is convenient and easy to implement, but there is a problem of discontinuous coding, which leads to the phenomenon that the ratio coding cannot match the fault type in practical applications [3]. At the same time, the three-ratio method is a method of judging the type of fault, which cannot play a role in early warning of transformer faults. Therefore, in actual use, the application effect of the characteristic gas three-ratio method is greatly restricted.

At present, domestic and foreign scholars have proposed a variety of improved transformer fault diagnosis methods based on dissolved gas analysis technology in oil. The key technologies used
involve support vector machines, neural networks, Bayesian networks, Raman spectroscopy, cloud theory, fuzzy theory, cluster analysis, etc., and have achieved good fault diagnosis results[4-8].

Therefore, this paper proposed a transformer condition evaluation method based on fault feature extraction. Based on the data of the key state quantity of dissolved gas in transformer oil, the transformer fault characteristics were extracted by various mathematical methods such as hypothesis test and logic language regression, and the operation state of transformer was judged by selecting new indexes which could reflect the operation state of transformer better.

2. Fault feature extraction

2.1. Frequency statistics and nonparametric test

In this paper, 62 indexes including single gas concentration, monthly standard growth rate of single gas concentration, ratio of two gases and relative content of multiple gases were selected. The frequency statistics was carried out by histogram with horizontal axis as statistical index value and vertical axis as statistical index. The concentration frequency distributions of H₂ and CH₄ were shown in Figure. 1 and Figure. 2.

(a) High temperature overheat fault transformer
(b) Normal transformer

Figure 1. Frequency distribution of H₂ concentration

(a) High temperature overheat fault transformer
(b) Normal transformer

Figure 2. Frequency distribution of CH₄ concentration

If a fault was found in the transformer detection at a certain time, or the fault was predicted by experts or judged as fault by three ratios, then the transformer was defined as fault from that moment (unless it was cleaned or repaired by oil, it was always determined as fault). Frequency statistics used two parts of data. The first part of the data was from the live detection data of 14 sets of 899 pieces of
data provided by a provincial power grid; the second part was from the live detection data of normal transformers provided by a provincial power grid, and a total of 531 data of 24 sets were selected.

The frequency distributions of the two parts of data were tested by nonparametric Mann Whitney Wilcoxon (MWW) test. The two opposite hypotheses of the test were:

\[ H_0: \text{the probability density function of the two parts of data was the same} \]

\[ H_1: \text{the probability density functions of the two parts of data were different} \]

The idea of the non-parametric Mann-Whitney-Wilcoxon (MWW) test was to merge the two parts of data into a whole, and then sort them from smallest to largest, and label them as 1, 2, ... from smallest to largest. This label was called rank. Calculated the sum of the ranks of the two parts of the data to be tested. If the probability density functions of the two parts of the data were the same, then the probability density distribution of the sum of the ranks could be obtained, so that the obtained rank sum could be rejected or accepted \( H_0 \) confidence level.

The results of MWW test showed that the probability density functions of the two parts of the 62 indicators were significantly different (the confidence of rejecting the null hypothesis was above 95%), but there were differences in the degree of significant difference. The higher the confidence of rejecting the null hypothesis, the more different the two parts of the data were. Therefore, we could use the confidence degree of rejecting the original hypothesis to measure the correlation between the index and the fault.

2.2. The inspection results

2.2.1. Index of high failure correlation

According to the above idea, the non-parametric Mann-Whitney Wilcoxon (MWW) test was used to obtain the confidence of rejecting the null hypothesis for each indicator. As shown in Table 1, the first 15 indicators were selected from the largest to the smallest according to the confidence of rejecting the null hypothesis of each indicator.

| Serial number | Indicators               | Serial number | Indicators               |
|---------------|--------------------------|---------------|--------------------------|
| 1 CH₄ concentration | 9 CH₄/(CH₄+C₂H₆+C₂H₄)       | 2 C₂H₆ concentration | 10 H₂/(H₂+C₂H₆+C₂H₄)       |
| 3 C₂H₄ concentration | 11 C₂H₆/(H₂+C₂H₆+C₂H₄)       | 4 CO₂ concentration | 12 C₂H₄/(H₂+C₂H₆+C₂H₄)       |
| 5 Total hydrocarbon concentration | 13 H₂/(H₂+C₂H₄+C₂H₂)       | 6 C₂H₆/H₂ | 14 C₂H₆/(H₂+C₂H₆+C₂H₂)       |
| 7 C₂H₆/H₂ | 15 C₂H₆/(H₂+CH₄+C₂H₆+C₂H₄+C₂H₂)       | 8 CO/CO₂        |                          |

2.2.2. Indicators of low failure correlation

According to the statistical results of frequency distribution and non-parametric MWW test results, there was little difference in frequency distribution of some indicators in the two parts of the data, such as CO concentration, CH₄/C₂H₆ ratio, CH₄/(CH₄+C₂H₆+C₂H₂) relative concentration and other indicators.

3. Index selection

3.1. Variable change

The above 15 variables related to the fault were denoted as X1-X15, and a threshold was taken for each variable, and X1-X15 was converted into 0,1 variables. The selection principle of the threshold was to make the difference between the fault transformer and the normal transformer at the threshold maximum, that is, the difference between the probability density of the fault transformer on the
threshold value and subtracting the probability density sum of the normal transformer on the threshold value was the maximum. As shown in the Table 2, 0 indicated that the transformer tends to be normal and 1 indicated that the transformer tends to fault.

| Serial number | Indicators | The value is 0 | The value is 1 |
|---------------|------------|----------------|----------------|
| X1            | CH₄ concentration | <=28.6          | >28.6          |
| X2            | C₂H₆ concentration | <=14.7          | >14.7          |
| X3            | C₂H₄ concentration | <=24.5          | >24.5          |
| X4            | CO₂ concentration | >=1519.1        | <1519.1        |
| X5            | Total hydrocarbon concentration | <=50.2          | >50.2          |
| X6            | C₃H₆/H₂       | <=0.887         | >0.887         |
| X7            | C₃H₆/H₂       | <=0.4           | >0.4           |
| X8            | CO/CO₂       | <=0.643         | >0.643         |
| X9            | CH₄/(CH₄+C₂H₆+C₂H₄) | >=0.5           | <0.5           |
| X10           | H₂/(H₂+C₂H₆+C₂H₄) | >=0.48          | <0.48          |
| X11           | C₂H₆/H₂       | <=0.355         | >0.355         |
| X12           | C₂H₄/(H₂+CH₄+C₂H₄) | <=0.31          | >0.31          |
| X13           | H₂/(H₂+C₂H₆+C₂H₄) | >=0.545         | <0.545         |
| X14           | C₂H₆/(H₂+C₂H₆+C₂H₄) | <=0.455         | >0.455         |
| X15           | C₂H₄/(H₂+CH₄+C₂H₆+C₂H₄+C₂H₂) | <=0.23          | >0.23          |

3.2. Logic language regression based on genetic algorithm
There were a total of 1430 pieces of data in the two parts. The first 9 pieces of data for every 10 pieces of data were 1287 pieces of data as the sample data, and the remaining 143 pieces of data were the test data. The transformer fault state was denoted as 1, the normal state as 0, and was represented by Y. The true value of transformer fault or not could be obtained by testing report or expert guess or three comparison value. Taking 15 variables 1 and 0 of X₁-X₁₅ as input, the regression was carried out with genetic algorithm through logic language and and or, and the frequency of using logic language was limited to less than 8 times (otherwise the relationship is too complex to have practical significance). The number of iterations was 200 and the number of children in each generation was 1000. The results were calculated by regression method.

The optimal regression equation in multiple regression were y = [(x₂ and x₉) or X₈] and [(x₈ or X₁₁) and X₁]. The specific judgment accuracy was shown in Table 3

| Accuracy of sample set judgment | The diagnosis result of the model was overheated equipment | The diagnosis result of the model was normal equipment | Judgment accuracy of prediction set | The diagnosis result of the model was overheated equipment | The diagnosis result of the model was normal equipment |
|---------------------------------|----------------------------------------------------------|------------------------------------------------------|-----------------------------------|----------------------------------------------------------|------------------------------------------------------|
| Equipment with actual overheating fault | 522/540 (96.7%) | 18/540 (3.3%) | Equipment with actual overheating fault | 56/60 (93.3%) | 4/60 (6.7%) |
| Equipment with actual health status | 29/747 (3.9%) | 718/747 (96.1%) | Equipment with actual health status | 4/83 (4.8%) | 79/83 (95.2%) |
As shown in the Table 3, it can be seen that the accuracy of judgment was basically stable in the sample set data and prediction set data, indicating that the results had strong stability and credibility. As shown in the Table 4, the regression results showed that 5 of the above 15 variable characteristics could be selected as the judgment index of whether the transformer was faulty.

Table 4. Five failure-related indicators screened out by logistic regression based on genetic algorithm

| Serial number | Indicators features |
|---------------|---------------------|
| X1            | CH₄ concentration  >28.6  |
| X2            | C₂H₆ concentration >14.7  |
| X₈            | CO/CO₂ >0.643  |
| X₉            | CH₄/(CH₄+C₂H₆+C₂H₄) <0.5  |
| X₁₁           | C₂H₄/(H₂+C₂H₆+C₂H₄) >0.355  |

4. Method comparison

According to the "Guidelines for the Analysis and Judgment of Dissolved Gases in Transformer Oil", the condition of using the three-ratio method is determined as the characteristic gas content exceeds the attention value. The comparison results between the method and the three ratios were shown in Table 5.

Table 5. Comparison of the accuracy of regression equation and three ratio method in population sample

| Overall judgment accuracy of logistic regression | The diagnosis result of the model was overheated equipment | The diagnosis result of the model was normal equipment | Overall judgment accuracy of three ratio method | The diagnosis result of the model was overheated equipment | The diagnosis result of the model was normal equipment |
|-------------------------------------------------|----------------------------------------------------------|-----------------------------------------------------|-------------------------------------------------|----------------------------------------------------------|-----------------------------------------------------|
| Equipment with actual overheating fault          | 578/600 (96.3%)                                          | 22/600 (3.7%)                                         | Equipment with actual overheating fault          | 504/600 (84.0%)                                          | 96/600 (16.0%)                                        |
| Equipment with actual health status              | 33/830 (4.0%)                                            | 797/830 (96.0%)                                       | Equipment with actual health status              | 31/830 (3.7%)                                            | 799/830 (96.3%)                                       |

Combined with the results of table 5, compared with the three ratio method, the method proposed by ontology has the following advantages:

1) Under the condition that the correct rate of normal transformer is basically unchanged, the correct rate of fault transformer judgment is greatly improved. This is because 62 candidate indicators are used in this method, and the scope of index consideration is much larger than that of three ratio method, and the influence of CO and CO₂ not considered by three ratio method on fault is considered in this method. With the help of the information of transformer operation state which is not used in the three ratios, the correct rate of normal transformer judgment is basically unchanged, and the correct rate of fault transformer judgment is improved.

2) There is only one ratio of two characteristic gases CO/ CO₂ in the new standard quantity of fault diagnosis. Compared with the three ratio method, the stability of fault diagnosis under low concentration is enhanced. When the concentration of the two characteristic gases is relatively low, it is easy to produce a larger measurement error than the true value. At this time, the ratio of the two characteristic gases will fluctuate greatly, and the relative concentration index will fluctuate...
less by the error. If there are two indicators, one is the ratio of the two characteristic gases: \( \frac{x}{y} \) and the other is the relative concentration: \( \frac{x}{x+y+z} \). Assuming that the measurement error changes to, the change rate of the ratio and relative concentration of the two characteristic gases is as follows:

\[
\begin{align*}
(kx / y - x / y) / (x / y) &= k - 1 > (kx / (kx + y + z) - x / (x + y + z)) / x / (x + y + z) \quad k > 1 \\
(kx / y - x / y) / (x / y) &= k - 1 < (kx / (kx + y + z) - x / (x + y + z)) / x / (x + y + z) \quad k < 1
\end{align*}
\]

The fluctuation of relative concentration index is less than that of the ratio of two characteristic gases.

3) No other external conditions are required. Relatively speaking, the three-ratio method is more of a fault type judgment method than a fault type judgment method. The precondition of using the three ratio method is that it has the characteristics of expert method to some extent when transformer is in fault by other ways. The method proposed in this paper uses regression method to find the index to judge the fault through screening a large number of indicators, so it does not need other external conditions.

5. Conclusion and Prospect
Hypothesis testing and genetic algorithm were used to perform regression analysis in this paper. A regression equation containing a higher correlation index with medium and high temperature overheating faults was obtained. The analysis results showed that, compared with the three ratio method, the average prediction accuracy of the proposed method was more than 95% under multiple cross validation, which had high engineering application value, and the stability of fault diagnosis under low concentration was enhanced. The next step is to build a high-dimensional linear model with variable points on this basis. The change point position of the transformer operating state index is calculated through the adaptive group cable estimation, and the position where the change point occurs is the early warning position of the transformer sub-health state, so as to realize the dynamic early warning of the transformer operating state.

Acknowledgments
This work is supported by Research on multi-spectral spectral video intelligent inspection technology for substation equipment under Science and technology project of State Grid Corporation of China (5200-201919048A-0-0-00).

References
[1] Guo, X., Song, Q., & Fan, X. (2013) Transformer fault diagnosis based on semi-supervised classifying method. Gaodianya Jishu/High Voltage Engineering, 39(5), 1096-1100.
[2] Xu Kun, Zhou Jianhua, Ru Qiushi, et al. (2005) Development and prospect of transformer oil dissolved gas on-line monitoring technology. High Voltage Engineering, 31(8): 30-32.
[3] Soualhi A, Clerc G, Razik H. (2013) Detection and Diagnosis of Faults in Induction Motor Using an Improved Artificial Ant Clustering Technique. IEEE Transactions on Industrial Electronics, 60(9): 4053-4062.
[4] Soualhi A, Clerc G, Razik H, et al. (2016) Hidden Markov Models for the Prediction of Impeding Faults. IEEE Transactions on Industrial Electronics, 63(5): 3271-3281.
[5] NILSSON J, BERTLING L. (2007) Maintenance Management of Wind Power Systems Using Condition Monitoring Systems Life Cycle Cost Analysis for Two Case Studies. IEEE Trans. On Energy Conversion, 22 (1):223-229.
[6] Firouzfar, M., Salah, P., & Madahi, S. S. K., (2010) Estimating the weight of main material for 63/20kV transformers with Artificial Neural Network (ANN). Power Engineering & Optimization Conference. IEEE.
[7] TIAN Bing. (2011) Multiple Linear Regression Analysis and Its Practical Applications. Yinshan Academic Journal (Natural Science Edition), 25(1):1 6-19
[8] MALIK H, JARIAL R K., (2011). Fuzzy-logic Applications in Cost Analysis of Transformer’s Main Material Weight. In: 2011 International Conference on Computational Intelligence and Communication Networks (CICN), Gwalior, India. pp. 386-389.