Fault diagnosis of power transformer based on tree ensemble model

Yunfei Liu¹, Jing Li¹*, Lin Qiao², Shuo Chen², Sai Liu⁴ and Jiahua Liu³

¹College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016 Jiangsu, China
²State Grid Liaoning Electric Power Supply Co, LTD, Shenyang 110004 Liaoning, China
³Nari Group Corporation/State Grid Electric Power Research Institute, Nanjing 211000 Jiangsu, China

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

Abstract. During the running operation of the oil-immersed transformer, some gases may be dissolved in the insulating oil which can be used to diagnose the incipient failure of the power transformer. This is the Dissolved Gas Analysis (DGA). This paper proposes a power transformer fault diagnosis method based on tree ensemble model (Extreme Gradient Boosting, XGBoost); constructing a large number of classification and regression trees (CART) to fit the residuals obtained by each learning. Compared with the commonly used SVM and BPNN methods, our method has a significant improvement in accuracy, F1-score, precision and recall.

1. Instruction

Power transformers are used to convert and transfer power generated by power plants to customers worldwide. Therefore, it is important to ensure that the power transformers are in good running condition so that they can provide reliable and continuous power, which is very necessary in the modern world [1]. Now, many power companies have implemented various running state assessment and maintenance measures on the running state of power transformers. Dissolved Gas Analysis (DGA) is one of them [2].

Transformers in normal state will decompose a very small amount of gases due to aging and cracking of insulating oil or solid insulation. These gases are mainly H₂, CH₄, C₂H₆, C₂H₄, C₂H₅, CO, CO₂ [3]. When the power transformer fails, the concentration of some gases will increase rapidly. However, gas concentration fluctuation range is large, which has a certain impact on the fault diagnosis of the power transformer. Therefore, in addition to the ordinary gas concentration, the ratios between different gases are also used to diagnose the fault of the power transformer. Commonly used ratios are IEC ratio (CH₄/H₂, C₂H₂/C₂H₄, C₂H₆/C₂H₅), Rogers ratio (CH₄/H₂, C₂H₂/C₂H₄, C₂H₆/C₂H₅), Dornenburg ratio (CH₄/H₂, C₂H₂/C₂H₄, C₂H₆/C₂H₅), which are all DGA methods [4][5]. DGA methods mentioned above can make estimation of the running state of power transformer, unfortunately, although they are simple to implement, they have problems such as incomplete coding and excessive boundaries. Besides, these methods often give different diagnosis results, indicating that these methods are very inaccurate [6]. These problems have driven many researchers to study methods based on machine learning to diagnose transformer faults. Support vector
machines (SVM) [1][7], neural networks(NN)[8][9], classification regression trees(CART) [10], principal component analysis(PCA) [7], fuzzy logic inference system(FLIS) [11] and other technologies [12] have been gradually applied to power transformer fault diagnosis and have achieved certain results. However, since transformer data is often very rare and difficult to collect, these intelligent algorithms, although able to achieve fairly good diagnosis results, tend to face problems such as overfitting.

XGBoost is an ensemble learning method based on regression trees that improves the traditional Gradient Boosting Decision Tree (GBDT). The excellent features of XGBoost drive us to explore the potential of XGBoost in power transformer fault diagnosis. The main contributions of this paper are as follows:

1. Our DGA data is collected from multiple sources, and the ratios between different gases are combined to form a relatively large DGA data set.

2. The use of ensemble learning (XGBoost) for transformer fault diagnosis and classification can allow the existence of a small number of missing values in the DGA data of transformer (this is very common due to the harsh running environment), Through the ensemble learning ability of XGBoost, a large number of classification and regression trees which can process highly oblique and polymorphic continuous data are constructed to diagnose the fault type of power transformer.

2. Extreme gradient boosting (XGBoost)

Extreme Gradient Boosting (XGBoost), a scalable machine learning system for tree boosting, is a novel classifier based on ensemble of classification and regression trees (CART) [13]. Before training, the data will be sorted firstly and then saved as a block structure. The subsequent iterations can use this structure repeatedly to greatly reduce the number of calculation and improve the classification speed of power transformers.

The data set with m-dimensional features can be represented as \( D = \{ (x_i, y_i) | |D| = n, x_i \in \mathbb{R}^m, y_i \in \mathbb{R} \} \), \( x_i \) means the features vector, \( y_i \) means the label of the fault type of power transformer. XGBoost is a tree ensemble model using K additive function to predict the output, and it can be defined as eq. (1).

\[
\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F
\]

is a collection of classification and regression trees (CART), \( q \) is a mapping function from sample data \( x_i \) to the leaf nodes of the CART, which is used to represent the structure of a tree, \( T \) indicates the number of leaf nodes in a tree. Each \( f_k \) is equivalent to a \( q \) mapping and the weight of its leaf nodes (score). This weight is a continuous value and contributes to an efficient optimization algorithm. \( w \) indicates the weight of the i-th leaf node.

In order to get the final XGBoost model, we need to train a set of regression trees. The objective function of this ensemble learning model is defined in eq. (2). \( l \) represents differentiable convex loss function, which is used to measure the difference between the real label \( y \) and the predicted label \( \hat{y} \). \( \Omega \) is a regular term that is used to smooth the learned weight and punish the complexity of the model to prevent overfitting.

\[
L(\phi) = \sum y l(y, \hat{y}) + \sum_k \Omega(f_k)
\]

\[
\Omega(f) = \gamma T + \frac{1}{2} ||w||^2
\]

This model is trained in an iterative manner. Let \( y_i^{(t)} \) be the prediction result of the i-th sample of the t-th iteration, the objective function is shown in eq. (3). \( f_t \) represents the new tree created by the t-th iteration. \( f_t \) is selected by eq. (3) to enhance the model. The residual between the result of previous iteration and real label is fitted by the new tree \( f_t \).
\[ L^{(t)} = \sum_{i=1}^{n} \left[ l(y_i, \hat{y}_i^{(t-1)} + f_r(x_i)) + \Omega(f_r) \right] \]

In the gradient boosting process, XGBoost uses a second-order Taylor expansion to optimize the objective function. The simplest form is shown in eq. (4), where \( g_i = \frac{\partial}{\partial y_i} l(y_i, \hat{y}_i^{(t-1)}) \) and \( h_i = \frac{\partial^2}{\partial y_i^2} l(y_i, \hat{y}_i^{(t-1)}) \), they are the first and second derivatives of the loss function respectively.

\[ L^{(t)} \approx \sum_{i=1}^{n} \left[ l(y_i, \hat{y}_i^{(t-1)} + g_i f_r(x_i) + \frac{1}{2} h_i f_r^2(x_i)) + \Omega(f_r) \right] \]

\[ \approx \sum_{j=1}^{T} \left[ \left( \sum_{i \in T_j} g_i \right) w_j + \frac{1}{2} \left( \sum_{i \in T_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \]

When the structure of a tree \( q(x) \) is given, the optimal weight of its leaf nodes can be calculated by eq. (5), the quality of the tree structure can be calculated by eq. (6).

\[ w^* = \frac{-\sum_{i \in T_j} g_i}{\sum_{i \in T_j} h_i + \lambda} \]

\[ \bar{L}^{(t)} = -\frac{1}{2} \sum_{j=1}^{T} \sum_{i \in T_j} g_i^2 w_j^2 + \gamma T \]

3. XGBoost-based power transformer fault diagnosis model

The method proposed in this paper is to use XGBoost to diagnosis the fault type of the power transformer based on the dissolved gas in the transformer insulating oil. For training, we use the data provided by NARI and combine some existing data from the network. Features and fault types of DGA data sets are shown in table 1.

| Table 1. Features and fault types of DGA data sets. |
|--------------------------------------------------------|
| DGA gas content (input) | Hydrogen(H2), Methane(CH4), Ethylene(C2H4), Ethane(C2H6), Acetylene(C2H2) |
| Ratios between different gases (input) | C2H2/C2H4, CH4/H2, C2H6, C2H2/(C1+C2), H2/(H2+C1+C2), C2H2/(C1+C2), CH4/(C1+C2), C2H6/(C1+C2), (CH4+C2H2)/(C1+C2) |
| Transformer fault category (output) | 0- Partial Discharge (PD) |
| | 1- Low energy discharge (D1) |
| | 2- High energy discharge (D2) |
| | 3- Thermal fault t < 300°C (T1) |
| | 4- Thermal fault 300°C < t < 700°C (T2) |
| | 5- Thermal fault t > 700°C (T3) |
| | 6- No Fault (NF) |
| | 7- Undefined (UD) |

We believe that transformer faults are related to the DGA gases and ratios between them. These 9 ratios are shown in table 1, in which \( C_1 \) and \( C_2 \) represent first-order and second-order hydrocarbons. \( C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9 \).

The DGA data of power transformer is input to XGBoost, and XGBoost builds the first CART according to the pre-processed data through the approximate algorithm in the split finding algorithm, and the predicted value is compared with the real value, then the residual is used as the new label information together with the sample data to construct the next CART tree to fit the residual.

The details of our fault diagnosis model for power transformer are as follows:

1. Input: The input data includes gas concentration, ratios and fault type.
2. Output: Set XGBoost to select softmax objective and return the diagnosed category by probability matrix.
3. Loss function: XGBoost only requires the loss function to be a differentiable convex loss function. We use mirror (multiclass classification error rate) and mlogloss (defined as the negative log-likelihood of the true labels given a probabilistic classifier's predictions) as our loss function.
4. Experiments and results

4.1. Data set and experiment setup

Due to the lack of existing DGA data sets, we collected and organized the data from multiple network resources. It may also brought data quality problems, but this is not our main consideration. Each sample in the DGA data consists of a series of concentration of dissolved gases, which can be seen in the table 2.

| H2  | CH₄ | C₂H₆ | C₂H₄ | C₂H₂ | C₂H₂/C₂H₄ | Fault |
|-----|-----|------|------|------|-----------|-------|
| 117 | 17  | 3    | 1    | 1    | 0.333     | ...... 0 |
| 595 | 32  | 18   | 4    | 65   | 3.611     | ...... 1 |
| 10  | 63  | 35   | 176  | 0.001| 0         | ...... 3 |

In order to eliminate the dimensional impact between features, data standardization is required to resolve the comparability between features. Through eq. (8), data in the DGA dataset is normalized to [0,1]. For entire experiment, we used a computer, Intel(R) Xeon(R) W-2102 CPU, 32G memory, python 3.6 and the operating system is Ubuntu 16.04.

4.2. Results and analysis

In this experiment, we tested the XGBoost classifier on a relatively large data set, and conduct in-depth analysis for them. We compare the XGBoost model with commonly used SVM and BPNN algorithms. 5-fold cross-validation is used to judge the pros and cons of the current model, and finally select the optimal XGBoost transformer fault diagnosis model.

Before training, we divided all the data into training set and testing set according to 7:3. The optimal parameters are max_depth=10, eta (learning rate)=0.22, gamma=0.18, subsample=0.69, colsample_bytree=0.56, min_child_weight =2, max_delta_step=10, the best number of regression trees required is 48. As the number of CART increases, merror and mlogloss rapidly decrease and stop at 48, then merror and mlogloss never again decrease significantly, and tend to be stable. Because the samples in the data set are not balanced, we not only evaluate the model from the accuracy rate, but also from precision, recall, and f1-score. The performance of optimal XGBoost model for each category is shown in the following table 3.

| Catagory | precision | recall | f1-score | samples |
|----------|-----------|--------|----------|---------|
| 0        | 0.75      | 0.41   | 0.53     | 29      |
| 1        | 0.60      | 0.69   | 0.64     | 52      |
| 2        | 0.89      | 0.85   | 0.87     | 93      |
| 3        | 0.81      | 0.73   | 0.77     | 70      |
| 4        | 0.88      | 0.91   | 0.90     | 261     |
| 5        | 0.87      | 0.56   | 0.68     | 93      |
| 6        | 0.79      | 0.81   | 0.80     | 137     |
| 7        | 0.90      | 0.97   | 0.93     | 450     |
| avg/total| 0.86      | 0.86   | 0.85     | 1185    |

It can be seen that in the categories with a large number of samples, the precision, recall, and f1-score scores are higher than those with fewer samples. For example, the UD category has the most samples, and the scores of the corresponding evaluation indicators are also the highest. Scores for categories with a less samples are less than expected. It means that imbalance and quality of the data set do have a certain impact on the results of the diagnosis model.

In the experiments of SVM and BPNN, the optimal weight parameters of BPNN are cross-validated to obtain. Grid search is also used for SVM. The parameters of the optimal SVM model obtained by grid search are \{C=1000, gamma=0.3, kernel="rbf"\}. In the experiment of BPNN, we also obtain the optimal model parameters by setting the grid search to compare the learning rate
and the number of layers and cells. The SVM model is not expected on our testing set, and several scores are quite low. BPNN performs slightly better than SVM on testing data set, but several evaluation indicators are still much lower than XGBoost.

In addition to comparing precision, recall, and f1-score, accuracy is also an important indicator. To some extent, it is the most important indicator. The average accuracy and top accuracy of the three models are shown in the following table 4.

| Model   | avg | top |
|---------|-----|-----|
| XGBoost | 0.88| 0.91|
| BPNN    | 0.71| 0.74|
| SVM     | 0.56| 0.62|

As can be seen from table 4, XGBoost is significantly better than SVM and BPNN in the four evaluation indicators of accuracy, precision, recall, and f1-score. The average accuracy is 0.17 higher than BPNN and 0.32 higher than SVM.

5. Conclusion
This paper proposes a fault diagnosis method for power transformer based on XGBoost (Tree Ensemble Model). The XGBoost-based transformer diagnosis model is compared with the commonly used SVM and BPNN methods and the experimental results show that the XGBoost-based transformer fault diagnosis model has a significant improvement in accuracy, F1-score and recall.

Acknowledgement
This research is 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)”.

Reference
[1] Z. Sahri, R. Yusof, Journal of Computer and Communications, 2, 22-31 (2014)
[2] A. Peimankar, S.J. Weddell, T. Jalal, et al. Swarm & Evolutionary Computation, 36, 62-75 (2017)
[3] Z. Sahri, R. Yusof, J. Watada, Industrial Informatics IEEE Transactions, 10(4):2093-2102(2014)
[4] A. Febriyanto, T.K. Saha. Power Engineering Conference. IEEE (2008)
[5] D.E.A. Mansour, Electrical Insulation & Dielectric Phenomena. IEEE (2012)
[6] R. Naresh, V. Sharma, M, Vashisth, IEEE Transactions on Power Delivery, 23(4):2017-24(2008)
[7] W. Shi, Y. Zhu, T. Huang, et al. Journal of Signal Processing Systems, 86(2-3):221-36 (2017)
[8] V. Miranda, A.R.G. Castro, IEEE Transactions on Power Delivery, 20(4):2509-16(2005)
[9] J.L. Guardado, J.L. Naredo, P. Moreno, et al. IEEE Power Engineering Review, 21(7):71-71(2007)
[10] X. Huang, L.I. Wenjunzi, S. Tong, et al. High Voltage Engineering (2016)
[11] S. Mofizul Islam, T.Wu, G. Ledwich. IEEE Transactions on Dielectrics and Electrical Insulation, 7(2):177-86(2000)
[12] S.S.M. Ghoneim, I.B.M. Taha, N.I. Elkalashy. IEEE Transactions on Dielectrics & Electrical Insulation, 23(3):1838-45(2016)
[13] T. Chen, C. Guestrin, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM (2016)