Fault Diagnosis and Asset Management of Power Transformer Using Adaptive Boost Machine Learning Algorithm

Suja Balaraman1, R. Madavan2*, S. Vedhanayaki3, S. Saroja4, M. Srinivasan5 and Albert Alexander Stonier5
1Department of Electrical and Electronics Engineering, Government College of Technology, Coimbatore
2Department of Electrical and Electronics Engineering, P.S.R. Engineering College, Sivakasi.
3Department of Electrical and Electronics Engineering, Akshaya College of Engg. & Technology, Coimbatore
4Department of Electrical and Electronics Engineering, MEPCO Schlenk Engineering College, Sivakasi.
5Department of Electrical and Electronics Engineering, Kongu Engineering College, Erode.
*Email: srmadavan@gmail.com

Abstract. Dissolved Gas Analysis (DGA) data of liquid insulation used to find the incipient faults such as partial discharge, thermal faults of various temperatures, discharge of high and low energy faults, combination of electrical and thermal faults in transformers. The conventional approaches of DGA namely Gas Ratio method, Duval triangle method and the Neural Network seems to be time consuming and sometimes yield erroneous results. In this paper, Adaptive BOOST machine learning algorithm is proposed, which is effective in classifying the transformer incipient faults. The results of proposed algorithm is compared with the results of different other machine learning algorithms such as K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree, Ensembler algorithm for the same set of transformers data. From the comparison, it is evident that ADABOOST machine learning algorithm performs well.

Keywords: Insulation Condition Monitoring, Fault Diagnosis, Dissolved Gas Analysis, Machine Learning Algorithms, ADABOOST Algorithm.

1. Introduction

Power transformers are one of the most expensive equipment in power system network which comprises of 60% of total power system investment [1]. Hence it is important to take care of power transformers in best way. In early decades the transformers maintenance is done through offline. In the last two decades, online condition monitoring is exercised. In online monitoring of transformers, the insulating oil in the transformer undergoes several chemical tests [2]. The dissolved gases in the oil are analyzed by gas chromatography method. Based on the dissolved gas concentration in oil, internal faults of the transformers like Partial Discharge (PD), Discharges of Low energy (D1), Discharges of High energy (D2), Thermal fault T < 300°C (T1), Thermal fault T > 700°C (T3), Thermal fault 300°C < T < 700°C (T2), Electrical and Thermal fault (DT) can be assessed [3]. The conventional approaches for interpreting dissolved gas level in insulating oil includes Doernenberg ratio, key gas ratio, Duval triangle, Rogers ratio and IEC 60599 methods and guidelines [4]. For the last two decades, Artificial Intelligence (AI) techniques comprising of Fuzzy Logic, Neural Network [5, 6], machine learning algorithms like wavelet network and support vector machine [6, 7] have been
widespread applied to diagnosis the condition of transformer insulation. AI technique uses characteristic data obtained from various transformers. AI determines correlation between the transformer insulation condition and oil characteristics and makes diagnosis on the transformer of interest. Further AI establishes complex and nonlinear relationships between DGA gases and transformer faults. However the above mentioned method has their inherent drawbacks. Fuzzy Logic Inference System suffers an accuracy loss and Neural Network suffers the problem of over fitting. In this work, ADABOOST algorithm is proposed for the classification of transformer incipient fault and based on the fault identified the transformer insulation condition is assessed.

2. Gas Generation Mechanism in Power Transformers
Mineral oil used as cooling and insulation purpose in transformers and it is the mixture of hydrocarbons. During the operating condition, with the application of thermal stress, chemical reaction takes place and it produces hydrocarbon molecules. Hydrogen (H2), carbon dioxide (CO2), carbon monoxide (CO), methane (CH4), ethane (C2H6), ethylene (C2H4) and acetylene (C2H2) are the gases generated during faults. The above said gases are also called as key gases. With the increase in gas content, more of it gets dissolved into the oil. Ultimately, a stage will come when the oil will be wholly saturated with the dissolved gases [9]. By assessing the rate of gas generation and the amount of generated gas present in oil, abnormality can be detected.

3. Dissolved Gas Analysis
During electrical and thermal stress, gases are formed due to decay of insulating oil [10]. When the major fault occurs, the concentration of gas production is higher and it can be detected by Buchholz relay. During the minor fault conditions, the amount of gas production is low and Buchholz relay takes time to sense the fault. Hence DGA is used to detect the transformer insulation condition during faults. The dissolved gases are detected using gas chromatography method and it’s in ppm [11]. Faults which are occurred in transformer along with the temperature are shown in Table 1. The conventional dissolved gas analysis methods given below.

3.1. Key gases ratio method
Key gas method is based on the quantity of individual fault gases released from the insulation oil during the occurrence of a fault which varies based on type of fault occurred. In this technique, individual gas concentration is accounted instead of the gas ratio for fault detection.

| Type of fault          | Indication |
|-----------------------|------------|
| Discharge of high energy (D2) | 0          |
| Discharge of low energy (D1) | 1          |
| Thermal and electrical faults (DT) | 2          |
| Partial discharge (PD) | 3          |
| Thermal fault, T < 300 °C (T1) | 4          |
| Thermal fault, 300 °C < T < 700 °C (T2) | 5          |
| Thermal fault, T > 700 °C (T3) | 6          |

3.2. Dorenberg and Rogers ratio method
In Dorenberg method, four important gas ratios namely CH4/H2, C2H2/C2H4, C2H2/CH4, and C2H2/C2H6 are considered for fault identification in transformers. Similarly, in Rogers method three gas ratio namely CH4/H2, C2H2/CH4, C2H2/C2H6 are considered for identifying the type of fault in the transformers [9].

3.3. Duval Triangle method
In this method concentration of methane, acetylene and ethylene are expressed in percentage and plotted as a dot in a triangular coordinate structure. This triangular chart is subdivided into different fault zones as shown in Figure 1 [12].
3

Where $A = (C_2H_2)$; $B = (C_2H_4)$; $C = (CH_4)$

4. Proposed Method for Fault Classification

In this work various machine learning algorithms are employed to classify the transformer incipient faults. The machine learning algorithm is defined as “the computer ability to learn without being explicitly programmed” [13]. The first evolved machine learning algorithm is neural network with back propagation algorithm and now-a-days there are different machine learning algorithms available namely Logistic Regression, Linear Regression, Decision Tree, SVM, K-Means, Naive Bayes, Random Forest, KNN, Ensembler etc. Further, based on the data available for learning the machine learning algorithms are classified into three types namely unsupervised learning, supervised learning and reinforcement learning.

4.1. Back Propagation Algorithm – Neural Network

The classification of fault is done by using Back Propagation-Neural Network. The three input neurons considered indicates the gases (% CH$_4$, % C$_2$H$_2$, % C$_2$H$_4$) and one output neuron for indicating the type of fault. Two networks are constructed different hidden layers. Network 1 is consisting of 5 hidden neurons and network 2 is with 10 hidden neurons.

The training samples for BPN network consist of Methane, Ethylene, and Acetylene in percentage as input data and type of fault as output. The training data for the algorithms are taken from the Duval triangle which includes 262 samples. The real time data of 131 power transformers across Tamil Nadu are taken for testing.

Percentage of Methane, Ethylene and Acetylene can be obtained by adding their concentration in ppm and dividing each gas by their total.

$$%C_2H_2 = \frac{100A}{A+B+C} \tag{3.1}$$

$$%C_2H_4 = \frac{100B}{A+B+C} \tag{3.2}$$

$$%CH_4 = \frac{100C}{A+B+C} \tag{3.3}$$

Figure1. Duval Triangle
Table 2. Classification of Fault using BPN

| Fault | Input Data | Network 1 (with single hidden layer) | Network 2 (with two hidden layers) |
|-------|------------|--------------------------------------|-------------------------------------|
|       |            | correct | incorrect | Accuracy (%) | correct | incorrect | Accuracy (%) |
| D2    | 10         | 5       | 5         | 50           | 7       | 30        | 70           |
| D1    | 10         | 7       | 3         | 70           | 6       | 4         | 60           |
| DT    | 16         | 5       | 9         | 31.25        | 15      | 1         | 93           |
| PD    | 1          | 0       | 1         | 0            | 0       | 1         | 0            |
| T1    | 7          | 5       | 2         | 71           | 5       | 2         | 71           |
| T2    | 19         | 4       | 15        | 21.05        | 16      | 3         | 84.2         |
| T3    | 27         | 19      | 8         | 70.37        | 25      | 2         | 92.59        |

By observing the Table 2 it is understood that the neural network with 10 neurons classifies the faults such as discharge of high energy, electrical and thermal fault, Thermal fault, $T > 700^\circ C$ effectively. However, Thermal fault with temperature range, $300^\circ C < T < 700^\circ C$ and $T < 300^\circ C$ (T1), discharge of high energy, partial discharge, are classified with lower classification accuracy.

4.2. Support Vector Machine

This is one of the machine learning algorithms that used to find an optimal separating hyperplane which maximizes the boundary of training. SVM supports binary classification, but to extend for multiclass radial and polynomial kernel function can be used. The kernel in SVM is defined as the dot product which can be described as

$$K(X, X_i) = \sum(X \cdot X_i) \quad (4.1)$$

Since in linear SVM the distance is a linear combination of the inputs dot product is used as the similarity measure, other than linear kernel for transforming the input space into higher dimensions various other kernels has been used such as a Polynomial Kernel and a Radial Kernel.

$$K(X, X_i) = X_i \cdot X_j \quad (4.2) \quad \text{linear SVM}$$

$$K(X, X_i) = (\gamma X_i \cdot X_j + C)^d \quad (4.3) \quad \text{polynomial SVM}$$

$$K(X, X_i) = e^{-\gamma \|X_i - X_j\|^2} \quad (4.4) \quad \text{RBF}$$

$$K(X, X_i) = \tanh(\gamma X_i \cdot X_j + C) \quad (4.5) \quad \text{sigmoid}$$

Here, $\gamma$ is an adjustable parameter of certain kernel functions.

4.2.1. One Vs One Classifier

It is a multi class strategy that consists in fitting one classifier per class pair. During prediction time, it requires to fit $n_{\text{classes}} \times (n_{\text{classes}} - 1) / 2$ classifiers, so the class which receives the most votes is selected. Because of $n_{\text{classes}}^2$ complexity it is usually slower than one vs rest method. However, this method seems to be advantageous for algorithms such as kernel algorithms which don’t scale well with $n_{\text{samples}}$. 


4.2.2. One Vs Rest Method

One vs Rest method is also known as One vs All or One against All strategy. In this strategy fitting one classifier per class occurs. For each classifier, the class is fitted against all the other classes where the fitting class is considered as positive and remaining all classes as negative. The main advantage of this method is its interpretability and because of that it is widely chosen.

Figure 2 illustrates the scatter plot of SVM, which helps to map two predictors at a time. The figure 2 is plotted between ethylene and methane gas.

![Figure 2. Scatter Plot of SVM](image)

Table 3. Confusion Matrix of SVM

| TRUE CLASS | D2     | D1     | DT     | PD     | T1     | T2     | T3     |
|------------|--------|--------|--------|--------|--------|--------|--------|
| D2         | 45     | 38     | 4      | 3      |        |        |        |
| D1         | 55     | 5      | 49     | 1      |        |        |        |
| DT         | 70     | 5      | 4      | 56     | 2      | 1      | 2      |
| PD         | 3      |        |        |        | 1      | 2      |        |
| T1         | 19     |        |        |        | 1      | 17     |        |
| T2         | 25     |        |        |        |        | 24     |        |
| T3         | 50     |        |        |        |        |        | 3      |

Table 3 shows the confusion matrix of SVM and it is constructed based on the fault classification of training data. For example while considering the D2 fault, 38 of the training data is correctly classified in respective fault but, 4 is misclassified as D1, 3 is misclassified as DT. Similarly for other faults it is summarized in table 3.

4.3. K-Nearest Neighbor Learning Algorithm

KNN algorithm can be applied for classification and regression process. It is a simple algorithm that stores all available cases and classifies new cases by a majority vote of its k neighbors. The objective being assigned to the class is common K nearest neighbors that can be measured by a distance function.

4.3.1. Parameter Selection

The best choice of k depends upon the input data given during the training. Generally, when larger values of k are used then the effect of noise on the classification gets reduced. But boundaries created between classes less distinct.
By observing the polar plot as shown in Figure 3, it is found that in KNN, misclassification of fault occur at large rate since the dotted line indicates majority of misclassification and also training time is high.

4.4. Ensembler Algorithm

Ensembler is a machine learning paradigm where multiple learners are trained to solve the same problem. In contrast to ordinary machine learning approaches which try to learn one hypothesis from training data, ensemble methods try to construct a set of hypotheses and combine them to use.

The occurrence of error in any model can classified into three components namely bias error, variance error and irreducible error.

\[
Err(x) = (e[f(x)]) - f(x))^2 + (E(f(X) - (E[f(x) | ]))^2 + \sigma \quad (4.6)
\]

\[
Err(x) = \text{bias}^2 + \text{variance}^2 + \text{irreducible error} \quad (4.7)
\]

Bias error is defined as calculating how much a predicted value is different from the actual value. Variance defined as how the prediction is made on same observation from each other. There are three techniques available to reduce these errors namely boosting, bagging, and stacking. Where boosting reduces variance error, bagging reduces bias error and stacking reduces either variance or bias error.

### Table 4. Confusion Matrix Using Adaptive Boost Algorithm

| Training Samples | TRUE CLASS | PREDICTOR CLASS |
|------------------|------------|-----------------|
| D2               | 45         | 36 3 6         |
| D1               | 55         | 3 48 4         |
| D1               | 70         | 3 4 62 1      |
| PD               | 3          | 3               |
| T1               | 19         | 1 19 3        |
| T2               | 25         | 2 24           |
| T3               | 50         | 1 2 1 43      |
| D2 D1 DT PD T1 T2 T3 |           |                |

4.5. Decision Tree

Decision tree is a type of supervised learning algorithm which is mostly used in classification problems. The major advantage of decision tree is that it works for both categorical and continuous input and output variables.
4.6. Adaptive Boost Algorithm

ADABOOST is a popular boosting technique which combines both Ensembler method and decision tree algorithm. It is said to be combining multiple “weak classifiers” into a single “strong classifier”. A weak classifier performs poorly, but on random guessing performs well.

\[ H(x) = sign(\sum_{t=1}^{T} \alpha_t h_t(x)) \]  

(4.8)

Here, the final classifier consists of ‘T’ number of weak classifiers. Where \( h_t(x) \) is the output of weak classifier’s’.

\( \alpha \) is the learning rate applied to classifier’s’ as determined by ADABOOST. In this paper learning rate is taken as 0.27.

From the Table 4 classification of fault using ADABOOST algorithm is observed and considering the PD fault, it is clear that even minority classes can classified effectively in this algorithm.

Table 5. Comparison of Fault Classification Using SVM, KNN, Ensembler, Decision Tree and Adaboost Algorithm

| Classifiers       | Method         | D2 | D1 | DT | PD | T1  | T2  | T3  |
|-------------------|----------------|----|----|----|----|-----|-----|-----|
| Cubic SVM         | 100            | 90 | 85 | 100| 85.7| 84.2| 92.59|
| Coarse KNN        | 70             | 60 | 75 | 0  | 28.5| 36.8| 48.14|
| Medium Tree       | 100            | 87 | 84 | 0  | 86  | 80  | 98  |
| Complex Tree      | 100            | 87 | 84 | 0  | 86  | 80  | 98  |
| Boosted Ensembler| 100            | 88 | 86 | 100| 79  | 83  | 98  |
| Bagged Ensembler  | 100            | 85 | 85 | 0  | 83  | 86  | 96  |
| ADABOOST          | 100            | 87 | 84 | 100| 86  | 80  | 98  |

Table 5 compares the accuracy level of SVM, KNN, Ensembler, Decision Tree and Adaboost Algorithms based on the faults. It clearly shows that, Adaboost algorithm have high accuracy compared to other fault classification algorithms. For some of the faults like D1, DT and T2, cubic SVM, Boosted Ensembler and Bagged Ensembler shows high accuracy than Adaboost algorithm. But in overall, Adaboost algorithm classifies the faults with higher accuracy than other algorithms.

Accuracy and training time of the algorithms are tabulated in Table 6. It clearly indicated that, Adaboost algorithm have higher accuracy and have lesser training time in fault classification.
Table 6. Comparisons of Attributes of Various Algorithms

| Attributes | Methods          | Accuracy (%) | Prediction Speed (Observation /Sec) | Training Time (Sec) |
|------------|------------------|--------------|-------------------------------------|---------------------|
|            | Cubic SVM        | 80.4         | ~3300                               | 2.2528              |
|            | Coarse KNN       | 44.5         | ~1300                               | 26.50               |
|            | Medium Tree      | 86.4         | ~8200                               | 17.63               |
|            | Complex Tree     | 86.4         | ~8200                               | 17.63               |
|            | Boosted Ensembler| 87.5         | ~280                                | 219.77              |
|            | Bagged Ensembler | 86.8         | ~180                                | 238.86              |
|            | ADABOOST         | 88.6         | ~3200                               | 1.6755              |

5. Results and Discussion

By observing the classification accuracy and attributes of SVM, it is found that the majority of classes are classified effectively and it take less time for training compared to neural network. However, over fit of minority classes occurs. While considering attributes and classification accuracy of KNN, it is found that misclassification of fault occurs at large rate and also it takes large training time. Both Decision tree and Ensembler have good classification accuracy but the training time is high. Adaptive Boost algorithm classifies the fault with higher accuracy and less training time. Further, over fit is eliminated for minority classes which are evident from confusion matrix. Once the fault is classified, transformer insulation condition is assessed by ranking method as shown in the Table 7 [11].

Table 7. Assessment of Insulation Condition

| Rating code  | Fault detected by ADABOOST algorithm                                      |
|--------------|--------------------------------------------------------------------------|
| 1-excellent  | Normal                                                                   |
| 2-very good  | Normal But High Total Combustible Gas                                    |
| 3-good       | Discharges of Low energy (or) Thermal fault T <300°C                      |
| 4-fair        | Thermal fault 300 °C < T <700 °C (or) partial discharge                  |
| 5-poor        | Discharges of Low energy (or) Thermal fault T >700 °C                    |

From Table 7, it is clear that the proposed ADABOOST algorithm classifies the incipient faults correctly and thereby assesses the insulation condition of transformers.

6. Conclusion

The prime aim of DGA based analysis is to identify the fault correctly from the gases generated during fault without de-energizing the transformer. The various machine learning algorithms such as SVM, KNN, DECISION TREE, ENSEMBLER and ADABOOST are discussed. From the analysis it is observed that ADABOOST algorithm classifies the incipient faults of the power transformer with high accuracy and less training time. The analysis is done by considering attributes such as accuracy, prediction speed and training time for same samples of training and testing data. Further, the over fit of minority classes are eliminated in this method. Also the insulation condition of the transformer is assessed based on the fault identified.

References

[1] Madavan R, Senthil Kumar S and Willuice Iruthayarajan M 2018 A comparative investigation on effects of nanoparticles on characteristics of natural esters-based nanofluids Colloids Surf. A Physicochem. Eng. Asp. 556 30–36
[2] Madavan Rengaraj, Saroja Subbaraj and Sujatha Balaraman 2017Multi-criteria decision-making methods for grading high-performance transformer oil with antioxidants under accelerated ageing conditions IET Gener. Transm. Distrib.11 16 4051 – 4058
[3] Shintemirov, Tang W and Wu Q H 2009 Power Transformer Fault Classification Based on Dissolved Gas Analysis by Implementing Bootstrap and Genetic Programming. IEEE Trans. Syst., Man, Cybern. Syst. 391 69–79

[4] Wang M, Vandermara A J, Srivastava K D 2002 Review of Condition Assessment of Power Transformers in Service. IEEE Electr. Insul. Mag. 18 612-25

[5] Wu H, Li X and Wu D 2011 RMP neural network based dissolved gas analyzer for fault diagnostic of oil filled electrical equipment. IEEE Trans. Dielectr. Electr. Insul. 18 2 495-498

[6] Yadaiah N, Ravi N 2011 Internal Fault detection technique for power transformers. Appl. Soft Comput. 11 5259-5269.

[7] Rengaraj Madavan and Subbaraj Saroja 2020 Decision making on the state of transformers based on insulation condition using AHP and TOPSIS methods. IET Sci., Measure. & Tech. 14 2137 – 145

[8] SUN Yan-jing, ZHANG Shen, MIAO Chang-xin, LI Jing-meng 2007 Improved BP Neural Network for Transformer Fault Diagnosis. J China Univ Min & Tech 17 1 138–142

[9] Huang Y C 2003A New Data Mining Approach to Dissolved Gas Analysis of Oil-Insulated Power Apparatus. IEEE Trans. Power Del. 18 4 1257-1261

[10] Madavan R and Sujatha Balaraman 2015 Performance analysis of transformer liquid insulation system under various environmental conditions. IEEE ConfProc, 83 – 86

[11] Chen W, Pan C, Yun Y X and Liu Y L 2008 Wavelet Networks in Power Transformers Diagnosis Using Dissolved Gas Analysis. IEEE Trans Power Del. 24 1 187-194

[12] Michel Duval 2002 A Review of Faults Detectable by Gas-in-Oil Analysis in Transformers. IEEE Elect Insul Maga 183

[13] Atefeh Dehghani Ashkezari, Hui Ma, Tapan K. Saha, and Chandima Ekanayake 2013 Application of Fuzzy Support Vector Maching for Determining the Health Index of the Insulation System of In-service Power Transformers. IEEE Trans. Dielectr. Electr. Insul 20 3

[14] A. A. Stonier, S. Murugesan, R. Samikannu, S. K. Venkatachary, S. Senthil Kumar and P. Arumugam, "Power Quality Improvement in Solar Fed Cascaded Multilevel Inverter With Output Voltage Regulation Techniques," in IEEE Access, vol. 8, pp. 178360-178371, 2020, doi: 10.1109/ACCESS.2020.3027784

[15] Sampath Kumar Venkatachary, Ravi Samikannu, Srinivasan Murugesan, Narasimha Rao Dasari, Ragupathy Uthandipalayam Subramaniyam, Nov. 2020, Economics and impact of recycling solar waste materials on the environment and health care, “Environmental Technology & Innovation”, Vol.20, no.101130, pp. 1-14.