

Machine Learning Approach to Dissolved Gas Analysis of Power Transformers

Kyle Deignan

University of Nevada Las Vegas Department of
Electrical and Computer Engineering
4505 S. Maryland Pkwy Box 454026
Las Vegas, NV 89154

Shahram Latifi

University of Nevada Las Vegas Department of
Electrical and Computer Engineering
4505 S. Maryland Pkwy Box 454026
Las Vegas, NV 89154

ABSTRACT

The purpose of this paper is to explore three different machine learning models to diagnose fault types inside power transformers. Power transformers are a key and expensive part of transmission and distribution systems. The traditional way of diagnosing faults has been to analyze the cooling oil for the presence of dissolved gases. From these gasses, methods of diagnoses have been developed such as Duval's triangle and the IEEE key gas concentration method. While these methods can be accurate in some cases, they do not consider the change in gas levels over time. This paper aims to develop models that use seven raw gas values over a period of five years to output a correct fault type for 12kV and above power transformers. Previous works have applied machine learning algorithms to a single sample of Dissolved Gas Analysis (DGA), however this paper aims to include multiple years of data. The rate of change of combustible gasses can be factored in when using multiple years of data, which is the key difference between this paper and previous work. Of the methods considered, it was found that Artificial Neural Networks (ANNs) had the best accuracy. An accuracy rate of 89% was achieved when using four hidden layers training on three years of data. K-Nearest Neighbors (K-NNs) and Support Vector Machine (SVM) also achieved adequate classification results with the best cases being 88% and 86% respectively. Previous work had an accuracy rate of 91% on a single sample of DGA classified using a single conventional method. An accuracy of 89% on multiple years of data is notable, because of the increased complexity of the training data and more in-depth method of classification.

General Terms

Dissolved Gas Analysis, Machine Learning, Power Systems.

Keywords

Artificial Neural Network, Dissolved Gas Analysis, K-Nearest Neighbors, Machine Learning, Power Transformer, Support Vector Machine

1. INTRODUCTION

Power transformers are often considered the most important devices within electrical power systems. They are expensive and crucial components for correct grid operation. Methods of identifying problematic transformers early are valuable as they often result in cost savings over the life of the transformer. One of the most useful methods of early transformer fault identification has become Dissolved Gas Analysis (DGA). As transformers are subjected to internal arcing or insulation degradation, dissolved gasses are released into the oil used for cooling. The early identification of faults types is achievable by analyzing the cooling oil for these dissolved gasses [1]. There are traditional methods that have

been developed, which often rely on certain gas ratios or thresholds to determine the fault type [2]. One example of conventional analysis is the Roger's Ratio Method. This method determines an internal fault type based off the combustible gas ratios present in the oil. Specifically, it considers the ratios of CH₄/H₂, C₂H₂/C₂H₄, and C₂H₄/C₂H₆ [1]. These ratios are compared with set thresholds that categorize the fault as thermal, energy discharge, or normal arcing. These thresholds allow several faults to be classified for a single sample, which is one problem with this method, as well as other conventional methods. While these methods prove to be effective, they only rely on one DGA sample and can often provide ambiguous results.

This paper intends to explore models that take into consideration a five-year spectrum of DGA data to diagnose the transformer. It is thought that the inclusion of historic data on a transformer will allow the trained models to consider the rate of gas change when classifying. The rate of change is becoming increasingly more important when evaluating DGA data. For example, a transformer with a small amount of acetylene would categorize as a faulty transformer when using many traditional forms of analysis. However, in practice, transformers with this condition tend to operate into their expected life if the amount of acetylene is not increasing over time. Looking at the spectrum of gases gives a more comprehensive view of the transformer, which is intended to be captured with the methods used in this paper.

1.1 Related Work

Machine learning applications have been applied to determine power transformer fault types in previous papers. Specifically, a hierarchal system of Support Vector Machines (SVMs) was developed based on the IEC method of DGA analysis [3],[4]. The first classifier determined if a fault was present, with the second classifier dividing the fault between thermal and electrical. This process continued until the fault was classified into one of the six IEC fault types. This hierarchical method achieved a total accuracy rate of 92% when tested, indicating that SVMs can provide high accuracy on a single DGA sample. Because of this high accuracy rate, SVM will be one of the algorithms explored in this paper.

Mirowski and LeCun have shown that Artificial Neural Networks (ANNs) can be implemented for DGA fault classification with accurate results [5]. Their method employed a neural network with one hidden layer using hyperbolic tangent as the activation function. Gradient descent was used to train their network, which resulted in an 89% accuracy rate when the test data was evaluated. Their study looked at SVM as well and determined that the use of a

Gaussian kernel resulted in the most accurate model. Various kernels will be explored in this paper to see if the results are in agreement.

Mirowski and LeCun also evaluated the use of k-Nearest Neighbors (K-NNs) for classification. This method provided the most accurate results in their paper at 91%. Because of its high accuracy, this method will also be explored in this paper. These previous works mentioned above used one DGA sample as the input to train the model. The key difference that this paper aims to explore is the use of multiple DGA samples to factor in the rate of change of combustible gasses to the fault classification.

1.2 Support Vector Machines

SVM are a set of supervised learning models in machine learning that can be used for outlier detection, regression, or classification [6]. Generally, SVM models are binary classifiers that use linear or nonlinear hyperplanes to separate two classes. By maximizing the margins around the hyperplane, the SVM model can make distinctions between the two classes. For multiclass classification, the One vs One algorithm or the One vs Rest algorithm can be used [7]. These are methods of extending the standard SVM to allow for more than two classes. For this study, a linear multiclass SVM will be utilized initially. Various kernel functions will also be tested to obtain the most accurate model. The python library sklearn will be used to develop the models.

1.3 Artificial Neural Networks

ANNs are another form of supervised learning that were inspired by the biological networks that make up the human brain [8]. They are a collection of nodes, or neurons, arranged into interconnected layers. Each connection has a weight that

training, the weights are adjusted to reduce the error found through backpropagation. Gradient descent is used during this backpropagation to update the weights to minimize error, resulting in a trained algorithm. Sklearn will also be used to develop this model.

1.4 K-Nearest Neighbors

K-NNs is a machine learning algorithm which can be used for both classification and regression [9]. This method works by assigning the output class of a sample according to the most common class among its nearest neighbors. The number of neighbors taken into consideration is user defined and changing this number will often result in different accuracy rates. If k is set equal to one, for example, the output class of a specific sample will simply be assigned the same class as the nearest data point. This method was found to be very accurate in classifying a single DGA sample and will be further explored in this paper for multiple samples [3].

2. MATERIALS AND METHODS

This paper uses three different models for fault identification into seven categories. These categories are normal condition, low thermal fault, medium thermal fault, high thermal fault, partial discharge, low energy discharge, and high energy discharge. These categories were chosen because the data that will be used has been classified into these seven categories by industry experts using a mix of traditional methods. The first method used will be the SVM. Because this is not a binary classification problem, the standard SVM model must be modified to handle multiple classes. This paper will use One vs One SVM. This is a method that constructs $k(k-1)/2$ classifiers where k is the number of classes. After the classifiers are constructed, a voting strategy is used to determine the correct classification for the given sample [6].

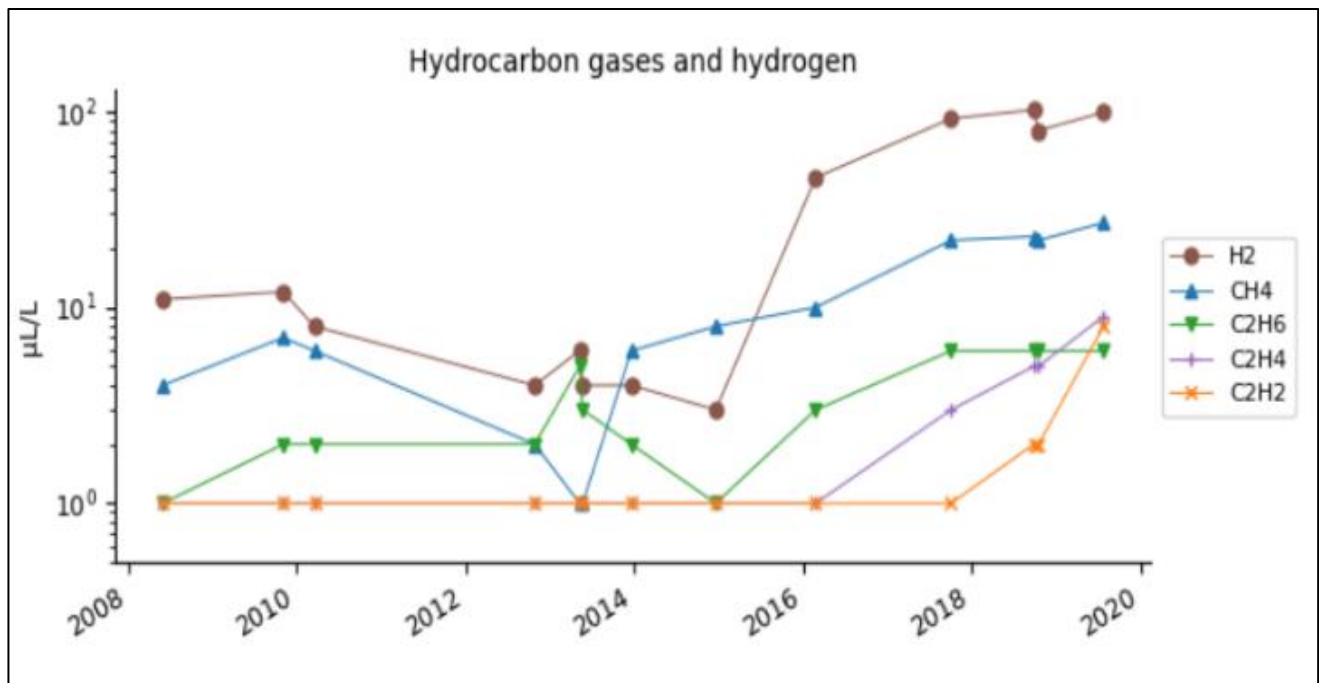


Figure 1 DGA Data

is applied to signals as they are fed through the network. An activation function is used to compute an output of each neuron based on all the inputs fed into that neuron. The signals pass through the network until the output layer is reached and a final classification is calculated. During

SVM has been applied to this problem in the past, but only on a single sample basis [5]. This paper will study the effectiveness of past and current gas values to determine fault conditions. The other two methods that will be explored are ANNs and K-NNs. All three of these methods have been

proven effective in classifying accurately when trained on a single sample which is why they are chosen for this paper [3],[5].

The input data that will be used comes from a local utility database, which currently has around 1200 transformers with historic data going back up to 20 years. This database includes transformers with voltages of 12kV or higher. A sample of the gas levels from this dataset are shown above in Figure 1. The five main gasses shown above will be used, as well as CO2 and CO, as they provide insight into the quality of the paper insulation [2]. The five main gasses were chosen because based off the traditional methods of analysis they provide the most insight into the type of fault that a transformer is experiencing. Essentially, each gas is generated at a certain temperature and those temperatures correlate with certain events that could be taking place. For example, acetylene is generated at 700° C, and this is most likely to occur when internal arcing is taking place, either in the coils or the tap changer [1].

The specific case shown in Figure 1 goes back to 2008; however, not all 1200 units contain a complete record like this case. Because of this, five years of data will be used, with a sample from every year. This was decided based on the quality of data available. It was also chosen to minimize the model’s complexity, as each new time value of data adds seven more dimensions to the feature vector. After extracting the DGA data and filtering for consistent five years of data, 379 transformers were found to train the model. Table 1 summarizes the input features, where the gasses will have five data points each for the five years. Based on the results found from using five years of data, multiple models will be trained to determine if increasing or decreasing the number of samples used for training provides a more accurate model.

Table 1 Features

Features						
H2	CH4	C2H2	C2H4	C2H6	CO2	CO

Table 2 shows the fault type classes associated with the DGA data. As mentioned, these classes were determined based on the data available. Given that the data was labeled by experts, using the most appropriate method for each situation, the trained model will reflect that expertise instead of recreating one specific method of analysis.

The models will be trained using the python library sklearn. This library offers all three of the methods that this paper is testing. Each algorithm has parameters that can be tuned. For

Table 2 Fault Type Classes

Fault Type Classes						
Normal	Low Thermal	Medium	High Thermal	Partial	Low Energy	High Energy
	Fault	Thermal Fault	Fault	Discharge	Discharge	Discharge

SVM, the kernel function will be changed and evaluated as well as the C parameter, which determines how much the model is affected by a misclassified point. In other words, a smaller C will result in a larger margin in the final model. This implementation has four different options for kernel functions: linear, polynomial, Gaussian radial basis function, or sigmoid. For this paper, linear, polynomial, and Gaussian will be used. Another interesting option available in this function is the ability to assign class weights when training. This essentially increases the penalty for misclassifying minority classes on unbalanced datasets [10]. This functionality will be tested and evaluated. For ANN, the parameters that will be changed and evaluated are the number of hidden layers, as well as the learning rate. Too large of a learning rate can result in convergence to a local minimum, so it is important to find an optimal rate. For K-NN, the number of neighbors taken into consideration will be varied in order to find the best accuracy rate. A range from one to 25 will be taken and the accuracy will be evaluated for each case to find the optimal k.

To summarize, the three different model types were trained with five, four, and three years of DGA data. For each case, the parameters outlined above were adjusted and evaluated to find a best case for each algorithm type. Table 3 below outlines the models and parameters that will be tested.

3. RESULTS

The DGA data was split 80/20 into training and testing data, respectively. The input data was first scaled using a function built-in to the library named “RobustScaler.” This function scales the data according to the quantile range, which defaults to the first and third quartile [10]. This is done for each feature separately. Once scaled, the data was split into training and testing data 80/20 respectively. All three algorithms were then implemented using three, four and five years of data. A simple accuracy function was used to evaluate each of the implementations. This function takes the number of correctly classified samples and divides it by the total samples to get an accuracy rate. The above parameters were varied and tested for each model case. Table 4 below gives a summary of the SVM results found.

The results above were found without using any weight balancing methods. Weight balancing was tried, and accuracy dropped much lower, so it was not pursued any further. Table 5 gives the results for ANN.

Finally, the results for K-NNs are shown below in Table 6. To reduce the table’s complexity, only intervals of five are shown below for the number of neighbors used for classification.

Table 3 Parameters

Model Type	Parameters	Parameter Values
SVM	C Value	0.1 - 0.5
	Kernel Function	Linear, Gaussian, and poly
	Weight Balancing	Balanced Based on Data
	Years of Data	3 through 5 years
ANN	Learning Rate	0.001, 0.01, 0.1
	Hidden Layers	1 - 5 layers
	Years of Data	3 through 5 years
K-NNs	K value	1 - 25
	Years of Data	3 through 5 years

Table 4 SVM Results

SVM	C Parameter				
	0.1	0.2	0.3	0.4	0.5
5 Years					
Linear	83%	86%	82%	82%	82%
Gaussian	54%	54%	58%	58%	66%
Poly	76%	78%	76%	76%	76%
4 Years					
Linear	78%	83%	82%	82%	82%
Gaussian	54%	55%	58%	66%	64%
Poly	71%	74%	72%	74%	74%
3 Years					
Linear	80%	80%	86%	84%	84%
Gaussian	54%	58%	66%	64%	64%
Poly	68%	71%	72%	72%	72%

Table 5 ANN Results

ANN	Number of Hidden Layers				
	1	2	3	4	5
5 Years					
Learning Rate = 0.001	82%	79%	78%	80%	82%
Learning Rate = 0.01	80%	83%	87%	80%	84%
Learning Rate = 0.1	76%	74%	72%	53%	68%
4 Years					
Learning Rate = 0.001	84%	86%	83%	86%	83%
Learning Rate = 0.01	84%	80%	83%	82%	84%
Learning Rate = 0.1	79%	61%	47%	76%	61%
3 Years					
Learning Rate = 0.001	83%	88%	84%	89%	87%
Learning Rate = 0.01	84%	83%	87%	83%	80%
Learning Rate = 0.1	78%	71%	71%	74%	74%

Table 6 KNN Results

K-NNs	Years of Data		
	3 Years	4 Years	5 Years
k = 1	78%	78%	76%
k = 5	88%	86%	84%
k = 10	87%	84%	83%
k = 15	82%	79%	78%
k = 25	75%	74%	72%

4. DISCUSSION

Of all the algorithms tested with this data, ANN yielded the best result. An accuracy rate of 89% was achieved when using four hidden layers with a learning rate of 0.001. The next best implementation found was K-NNs, when using five neighbors to classify, which had a slightly lower accuracy, at 88%. SVM had mixed results depending on the parameters but was able to achieve an 86% accuracy rate when a linear kernel was used with three years of data.

In general, it can be seen that using more years of data resulted in a lower accuracy rate. This is thought to be because the feature vector becomes too complex for the model to accurately fit to, as each new year of data adds seven new gas values to the feature vector. Interestingly, when using SVM, the linear kernel resulted in a much better model than the Gaussian or polynomial kernels, which goes against what was found previously by Mirowski and LeCun [5]. This may be a result of using multiple samples to train, where as their study was aimed at correctly classifying a single DGA sample. Similar accuracy was found as the previous works, outlined above. This is notable because of the increased complexity of the feature vector for this study and a more in-depth method of classification.

5. CONCLUSIONS

Three different machine learning algorithms were explored to find the validity of applying them to a five-year spectrum of DGA samples. A database of 379 transformers was used to train the models. These transformers had been classified into different fault types by industry experts, using a mix of conventional DGA methods. After tuning parameters to achieve the highest level of accuracy, it was found that Artificial Neural Networks produced the best results, with an accuracy rate of 89%. This is comparable to previous works which only considered one DGA sample classified off a single conventional method, such as the IEC fault identification method. K-NNs and SVM also achieved comparable accuracy rates at 88% and 86%, respectively. In all, it was found that these machine learning algorithms can be utilized to classify fault types within transformers based on multiple years of DGA samples

6. ACKNOWLEDGEMENTS

We would like to thank Meagan Madariaga-Hopkins for proofreading of this paper

7. REFERENCES

- [1] Hamrick, L.: Dissolved gas analysis for transformers. NETA World Journal. 1-3 (2009)
- [2] IEEE Std. C57.104-2008.: Guide for the interpretation of gases generated in oil-immersed transformers. (2008)
- [3] J. Singh, P. Kumari, K. Kaur and A. K. Swami, "Condition assessment of power transformer using SVM based on DGA," 2016 Al-Sadeq International Conference on Multidisciplinary in IT and Communication Science and Applications (AIC-MITCSA), Baghdad, 2016, pp. 1-5, doi: 10.1109/AIC-MITCSA.2016.7759957
- [4] F. R. Souza and B. Ramachandran, "Dissolved gas analysis to identify faults and improve reliability in transformers using support vector machines," 2016 Clemson University Power Systems Conference (PSC), Clemson, SC, 2016, pp. 1-4, doi: 10.1109/PSC.2016.7462827
- [5] P. Mirowski and Y. LeCun, "Statistical Machine Learning and Dissolved Gas Analysis: A Review," in IEEE Transactions on Power Delivery, vol. 27, no. 4, pp. 1791-1799, Oct. 2012, doi: 10.1109/TPWRD.2012.2197868.
- [6] Rathgamage Don, Duleep Prasanna W., "Multiclass Classification Using Support Vector Machines" (2018). Electronic Theses and Dissertations. 1845
- [7] Burges, C.: A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery. 2, 121-167 (1998).
- [8] Zupan, J.: Introduction to Artificial Neural Network (ANN) Methods: What They Are and How to Use Them. Acta Chimica Slovenica. 1994
- [9] Okfalisa, I. Gazalba, Mustakim and N. G. I. Reza, "Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification," 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, 2017, pp. 294-298, doi: 10.1109/ICITISEE.2017.8285514.
- [10] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.