

Feature Engineering and Classification Models for Partial Discharge Events in Power Transformers

Jonathan Wang
Rice University
jw96@rice.edu

Kesheng Wu, Alex Sim
Lawrence Berkeley National
Laboratory
kwu@lbl.gov, asim@lbl.gov

Seongwook Hwangbo
Hyundai Electric & Energy Systems
Co., Ltd.
smhwangbo@hyundai-electric.com

ABSTRACT

To ensure the reliability of power transformers, they are monitored for partial discharge (PD) events, which are symptoms of transformer failure. Our goal is to classify PDs to gain an understanding of the location of failure. We develop a small set of features and a stacking ensemble that outperform larger feature sets and other models in both accuracy and variance.

1 INTRODUCTION

Power transformers are a key element of electric power infrastructures. While they have become more reliable, transformers are still susceptible to failure. This work focuses on analyzing partial discharge (PD), an internal arcing event that signals transformer failure.

Certain types of PDs are correlated with different parts of the transformer. Therefore, determining the type of PD provides a rough location for the PD source. We can then install UHF sensors around the rough position to collect PD signals to identify the precise position of the PD for repair.

We present a small set of features that can efficiently represent the PD data as well as a stacking ensemble model that can classify PD types more accurately than existing methods. Our feature set attains 99.31% accuracy with a Random Forest classifier, the best tested single model. This accuracy is improved to 99.61% with our stacking ensemble, which also achieves half the variance of single models.

2 METHODS

The goal of our work is to classify four types of PD - *corona*, *floating*, *particle*, and *void*. We accomplish this by:

- Extracting features from signal data
- Training machine learning model on features

Our data is 328 PD signals gathered by the transformer sensors labelled as 85 *corona*, 99 *floating*, 80 *particle*, and 64 *void*. Each data sample contains 3840 magnitude points over one second. These points are broken up into 60 cycles of 64 phases. Figure 1 shows heatmaps of each type of PD. The x and y axes indicate the phase and cycle and the color indicates the magnitude at that time.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

BDCAT'17, , December 5–8, 2017, Austin, Texas, USA
© 2017 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-5549-0/17/12.
<https://doi.org/10.1145/3148055.3149270>

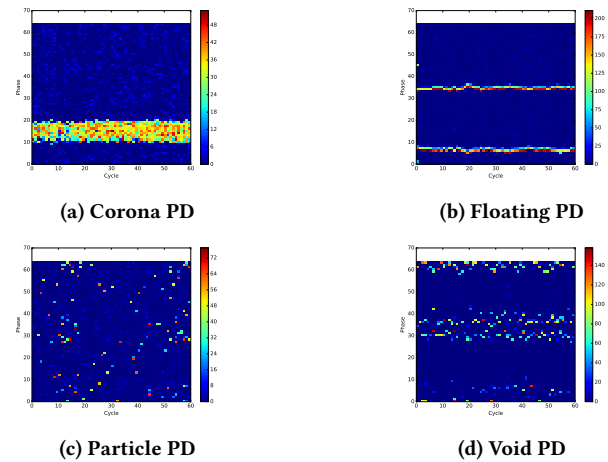


Figure 1: Heatmaps of Sample PDs

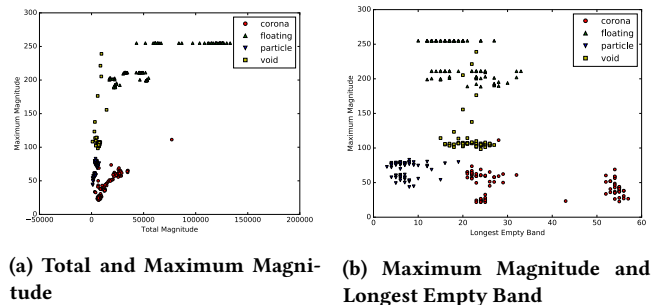


Figure 2: PD Samples Plotted along Features

Since using all 3840 magnitude points as features would be too costly to model, we want to extract a smaller set of features from the signals.

The set of features that we examine is:

- the maximum magnitude out of all 3840 points
- the total magnitude of all 3840 points
- the length of the largest empty phase band

The **length of the largest empty band** quantifies the distribution of the signal. We define an empty band as consecutive phases without significant magnitude (40% of maximum magnitude).

These features are plotted with the PD type in Figure 2, which shows that the data points are almost separable based on these three features.

Table 1: Prediction Accuracy \pm Standard Deviation by PD Type

Classification Method	PD Type				Total
	Corona	Floating	Particle	Void	
SVM	0.9915 \pm 0.014	1 \pm 0	0.9954 \pm 0.014	0.9789 \pm 0.042	0.9923 \pm 0.010
Logistic Regression	0.9997 \pm 0.002	0.9882 \pm 0.024	0.9680 \pm 0.035	0.9809 \pm 0.024	0.9847 \pm 0.011
Random Forest	0.9905 \pm 0.014	1 \pm 0	0.9954 \pm 0.012	0.9832 \pm 0.035	0.9931 \pm 0.009
Gradient Boosting	0.9672 \pm 0.030	1 \pm 0	0.9862 \pm 0.024	0.9785 \pm 0.035	0.9838 \pm 0.012
Fuzzy SVM (FSVM)	0.9859 \pm 0.023	1 \pm 0	0.9943 \pm 0.017	0.9712 \pm 0.029	0.9893 \pm 0.011
Best Stacking Model	0.9985 \pm 0.007	1 \pm 0	0.9984 \pm 0.008	0.9836 \pm 0.021	0.9961 \pm 0.005

We test several classification methods, including Gradient Boosting, Random Forest, Logistic Regression, SVM (Support Vector Machine) and FSVM (Fuzzy Support Vector Machine). For each of our experiments, we split our dataset into training and validation sets (60:40). We train the models on the training set and score prediction accuracy based on the validation set. This process is repeated 100 times for each model and the results are averaged to achieve more consistent results.

Certain models predict certain PD types better. To take advantage of the strengths of each model, we implement a stacking ensemble classifier, which is comprised of two levels of classification. We train a set of level one classifiers as in our previous experiments. The outputs of these classifiers are then used as features for the level two classifier. To build an optimized model, we test several parameters, such as what classification methods to use in each level and what features to extract from the level one classifiers, as well as whether to use the probabilities or prediction from the level one classifiers and whether to include the original features in the stacking model.

3 EXPERIMENTAL RESULTS

Table 1 compares the performance of each model for each PD type. We can see that SVM, FSVM, and Random Forest are fairly strong classifiers overall. However, Logistic Regression performs better in classifying only *corona* PDs. In addition, over multiple experiments, SVM and FSVM occasionally misclassify *floating* PDs, although this happens very rarely and is not represented in the table. In contrast, Random Forest and Gradient Boosting have no variance in the classification of *floating* PDs.

Due to the unique strengths of the various models, we combine them into an ensemble using stacking. We try several parameter variations of the stacking classifier, but our selection is based on the observation that Logistic Regression has much better *corona* classification scores, and Random Forest and Gradient Boosting have more consistent *floating* classification scores. The parameters are the classifiers to stack, the model to stack with (meta-classifier), whether to use the model probabilities as features, and whether to include the original features in the stacked model. The classifiers that we consider are Gradient Boosting (GB), SVM, Logistic Regression (LR), and Random Forest (RF).

The resulting stacking model uses 2 SVMs, Logistic Regression, and Random Forest provides the best results. This classifier is shown in Figure 3. Although it does not achieve the maximum *corona* accuracy of Logistic Regression, it offers the best total accuracy.

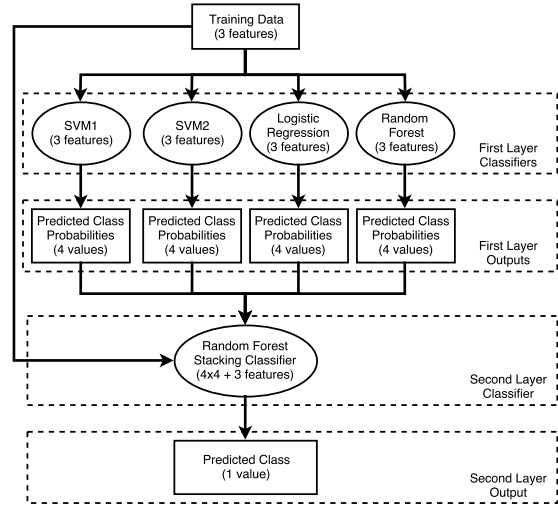


Figure 3: Diagram of Final Stacking Classifier

This model outperforms any single classification model in terms of prediction accuracy with a total accuracy of 0.9961 whereas Random Forest, the best single model, has an accuracy of 0.9931. For each PD type, the stacking model outperforms Random Forest, particularly for *corona* PDs, where Random Forest scores 0.9905 compared to the 0.9985 of the stacking model. The stacking ensemble also reduces the variance of the total accuracy by half and obtains the best variance for each PD type. The total variance for each single model hovers around 0.010, while the stacking model only has a total variance of 0.005.

4 CONCLUSION

As a result of our work, we determine an efficient feature set that can be used to accurately classify PD events. We also present a stacking ensemble strategy that combines the strengths of several prediction models to outperform existing classification methods, consistently achieving more than 99% accuracy with low variance.

ACKNOWLEDGMENT

This work was supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.