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

Jonathan Wang[1], Kesheng Wu[2], Alex Sim[2], Seongwook Hwangbo[3]
[1]Rice University, [2]Lawrence Berkeley National Laboratory, [3]Hyundai Electric & Energy Systems Co., Ltd.

U.S. DEPARTMENT OF ENERGY
Office of Science

RESEARCH GOAL

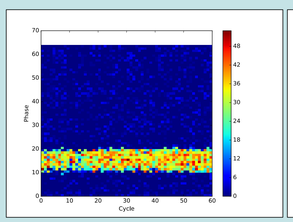
To design an efficient feature set and prediction model to classify different types of partial discharge based on signal data

BACKGROUND

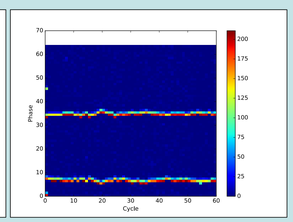
- Insulation failure is the most common type of transformer breakdown
- **Partial Discharge (PD)**, an internal arcing event, is a symptom of insulation failure
- Different types of PD correlated with certain areas of transformer
- Determine PD type to find rough location
  - **Extract features from signal data**
  - **Train classification models on features**
- Install ultra-high frequency sensors around rough location to identify precise PD origin

DATA

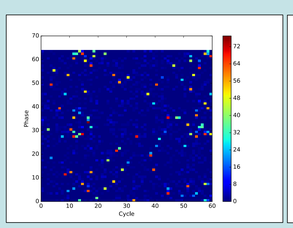
- Samples contain 3840 data points divided into 60 cycles of 64 phases
  - 328 total samples
  - Unique patterns for each PD type



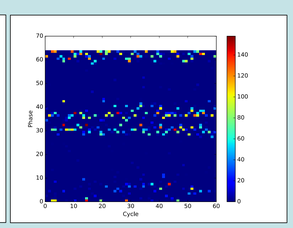
(a) Corona PD (85)



(b) Floating PD (99)



(c) Particle PD (80)



(d) Void PD (64)

Figure 1: Signal data of the 4 PD types we examine and the number of samples of each type

FEATURE ENGINEERING

- Extract smaller feature set from 3840 points of PD signal
- Each meta-feature represents entire data sample
  - **Maximum magnitude** out of 3840 points
    - Averaged over top 3 points to account for outlier points
  - **Total magnitude** of 3840 points
    - Represents average magnitude
  - **Length of longest empty phase band**
    - Estimate of signal distribution

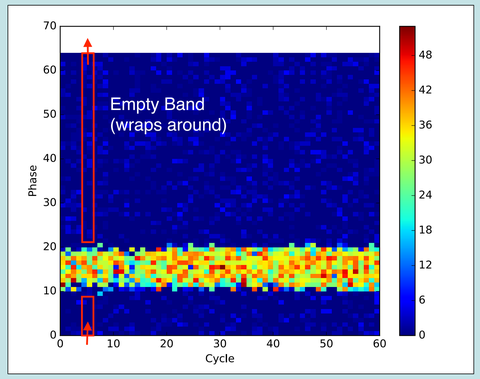


Figure 2: Empty phase band is a continuous set of phases with no magnitudes greater than 40% of maximum magnitude.

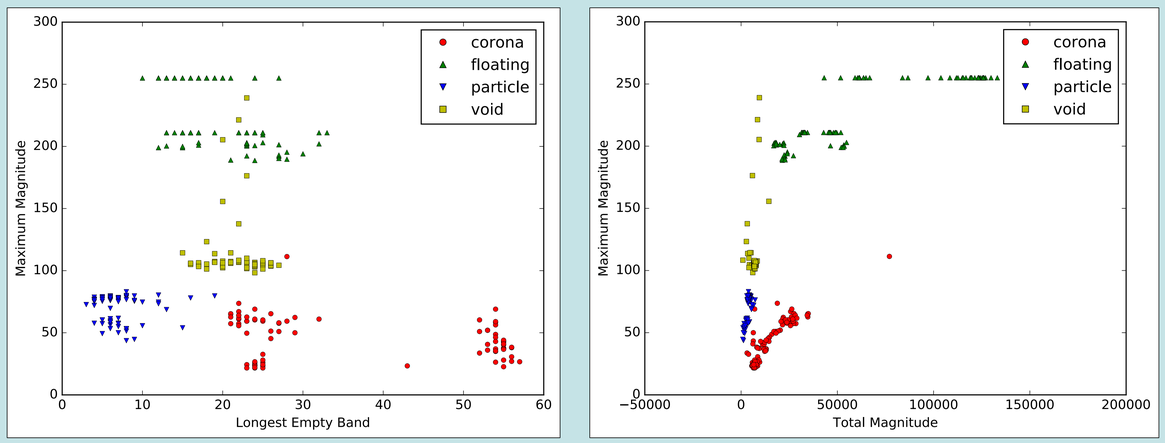


Figure 3: The PD data is almost separable based on the 3 features. There are fairly distinct clusters with few outliers, especially when considering all 3 features together.

CLASSIFICATION

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

Figure 4: SVM (Support Vector Machine), FSVM, and Random Forest have high overall accuracy. However, Logistic Regression performs best for just Corona PDs. The stacking ensemble model has the best overall accuracy with lower variance than any single model.

PD Type	Recall	Precision
Corona	0.998	1
Floating	1	0.991
Particle	0.999	1
Void	0.988	1

Figure 5: The precision and recall for all PD types are very close. However, we notice that the most common errors are void PDs being misclassified as floating PDs.

STACKING

- Several successful classification models
  - Certain models perform better for specific PD type
  - Combine strengths of each model
- Stacking ensemble
  - Train multiple first layer classifiers
  - Use outputs as input to second layer classifier

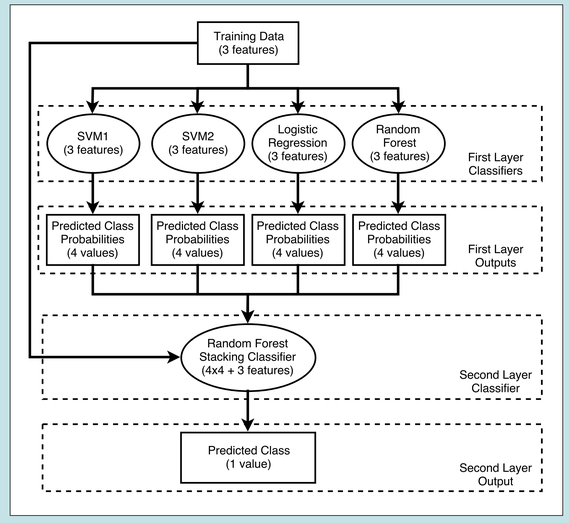


Figure 6: Diagram of stacking classifier. Parameters to test are first and second layer classifiers as well as how to translate first layer outputs to second layer input.

CONCLUSION

- Define feature set that efficiently represents PD signal samples
  - Better accuracy and less data than phase based features
- Stacking ensemble outperforms any single model
  - Higher accuracy and lower variance

ACKNOWLEDGMENTS

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

CONTACT

- Jonathan Wang
- Rice University
- Email: jw96@rice.edu