Study Committee D2

Information Systems and Telecommunication

Paper 10849_2022

DATA AUGMENTATION FOR CLASSIFICATION OF THE PARTIAL DISCHARGE PATTERN CONSIDERING IMBALANCE AND PHASE UNCERTAINTY OF THE TRAINING DATASET

G. HWANG*, H. KWON, S. KOO

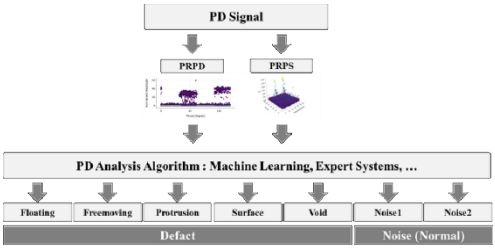
LS ELECTRIC Co., Ltd

Motivation

- Partial discharge (PD) analysis is typical classification problem, and accordingly, machine learning (ML) can exhibit high performance when there is a lot of data. The more types of data augmentation techniques could help further improvement of the model performance.

Method/Approach

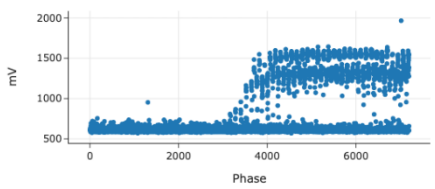
- PD data is mainly acquired from ultra-high frequency (UHF) sensors which are installed in gas insulated switchgear (GIS) and UHV transformer. A strong pulse is generated by a specific phase depending on the type of insulation defect.
- A unique pattern is shown when converting through PRPS or PRPD method after filtering noise in the data preprocessing stage. The PD diagnosis is a method to detect and classify defects with types using these properties.



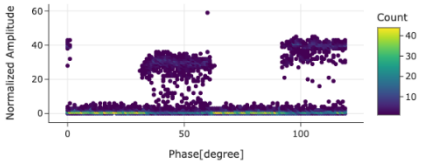
- To resolve the data uncertainty caused by phase shift and data imbalance, this paper presented the phase into a total of four cases, ±60° and ±120°, and moved the phase of the existing data according to each case.
- Data was generated by adding a noise to the data assuming various noises that may be generated by the external environment.

Data acquiring and processing

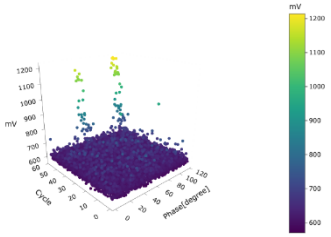
- Ultra-high frequency (UHF) are used for PD measurement.



- The PD has a constant pattern depending on phase, which is different from a noise and it can be distinguished by a cause of the PD generation.
- The PRPD is a method of converting signals using phase, pulse magnitude, and pulse frequency.



- The PRPS uses the same phi-q-n method as PRPD, but its data is presented on a time axis.



Issues of the processed data

- Data imbalance between abnormal data and normal data. UHV devices such as GIS and Tr. are highly reliable because they could be critical impact on the system at the time of failure.
- It is very difficult to check the labeled defect data because the observed PD defect pattern must be checked after opening the seal of the insulator to accurately indicate the location of the physical insulation problem.
- The label has a hierarchical structure. A target value that the model should predict is classified into a defective PD and a normal noise. The PDs are again classified according to the location of defects such as void, protection, freemoving, and floating, and even the same defects can be distinguished to several types again.
- Since each axis of processed data by PRPD/PRPS has different physical meanings such as phase, cycle, and amplitude, this should be considered when selecting kernel of CNN-based AI algorithm.
- Since the sensor collects PD signals without distinguishing between phases a, b, and c, it may be collected by moving ±60° and ±120° from the actual phase.

Study Committee D2

Information Systems and Telecommunication

Paper 10849_2022

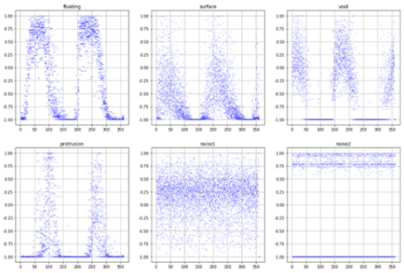
DATA AUGMENTATION FOR CLASSIFICATION OF THE PARTIAL DISCHARGE PATTERN CONSIDERING IMBALANCE AND PHASE UNCERTAINTY OF THE TRAINING DATASET continued

Data augmentation for PD pattern

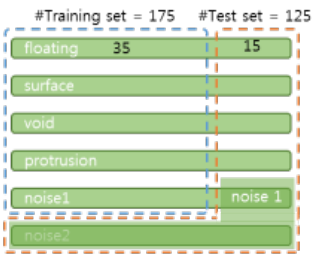
- Data filtering for removing noise is required to improve the classification performance of the data augmentation.
- After filtering, data preprocessing by using PRPD/PRPS is performed.
- Data augmentation is performed by moving the phase of the existing data according to four cases, ±60° and ±120°.

Experimental setup & test results

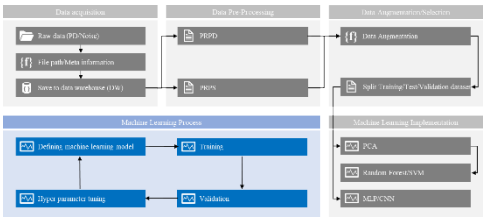
- Use data obtained from 170kV GIS currently in commercial operation.
- Defects are distinguished by four types: floating, surface, void, and protusion, and normal data are separated by noise 1 and noise 2 which are different types.



- The data of each class were splited into train data and test data at a ratio of 7:3.
- In the case of noise 2, it was excluded from learning and used only as test data. The reason is that all noise patterns cannot be learned in the model because the noise pattern appearing in the field is not fixed but continues to change depending on the surrounding environment.



- Workflow of training macine learning model.



- Four cases were tested.
  - A simple multi-layer perceptron model(MLP) with only one hidden layer.
  - The simple MLP with one hidden layer + Augmented data from 175 to 10,000 by adding white noise to the existing waveform.
  - The simple MLP with two hidden layer + Augmented data.
  - ResNet based Convolutional Neural Network(CNN) + Augmented data.
- Result of the cases.

Model	Accuracy [%]
MLP (1 hidden layer)	87.2
MLP (1 hidden layer) + Data Augmentation	92.0
MLP (2 hidden layer) + Data Augmentation	96.0
CNN + Data Augmentation	100.0

Conclusion

- The type of model and the size of the dataset have a great influence on the accuracy of PD diagnosis model.
- It is very important to understand the characteristics of PDs in tuning various parameters such as kernal size, pooling method, etc.
- Data augmentation techniques are one of the methods to improve the performance of models in situations where it is difficult to collect failure data.