# DATA MINING METHODS TO PREDICT FAILURE DUE TO PARTIAL DISCHARGE

Gareth POULTON
IPEC Ltd – UK
gpoulton@gmail.com

Colin SMITH
IPEC Ltd – UK
colin@ipec.co.uk

Wai-Shing HO
University of Hong Kong - HK
wsho@cs.hku.hk

## *ABSTRACT*

*A current weakness in the use of Partial Discharge (PD) analysis for automated failure prediction is the difficulty in attributing accurate time-to-failure or confidence levels to any given measurement.*

*We have attempted to address this through the use of several data mining methods on a large database of both failed and healthy measurements taken over the course of 2-5 years on almost a thousand assets. These methods and their relative usefulness are discussed.*

## OVERVIEW

Prediction of likely failure can be seen as a classification problem, in which a channel (i.e. the data from a sensor attached to a cable or switchgear) may be classed 'good' (healthy) or 'bad' (likely to fail). Working backwards from known 'good' and 'bad' channels - that is to say, channels whose eventual good health or failure is known - we can construct rule sets or decision trees that predict this outcome using well-documented pattern classification methods.

The authors are not aware of any previous work using this method to classify PD signals, however other strategies have been used in the analysis of PD data [1][2][3].

### About the data set

The data used was gathered from just over a thousand circuits in EDF Energy substations around London over the course of up to five years. It has been the subject of past CIRED papers including Walton et al [4] and Smith et al [5], both of which describe the acquisition process in more detail.

### Preparation

The majority of classification methods are not immediately well-suited to time series data. As such, various methods of summarizing the characteristics of an asset in a single number were rated for usefulness. These involved taking such things as mean, maximum, standard deviation, difference between earlier and later means, and so forth, each across several different time periods, and for various methods of measuring PD levels.

The bulk of the classifiers were eventually constructed using four-week periods, a figure chosen based on manual inspection of the data. Thus, each four-week period from each channel was considered as an independent data point with its own label of 'good' or 'bad', and each had several methods of summarizing such features as:

- The positive and negative peak mV and area of the largest individual discharge in an AC cycle (area being directly proportional to pC charge)
- A count of the number of discharges that happen in a cycle, for each of 16 thresholds.
- the total sum of the areas of all detected discharges (pC)
- total number of discharges in a cycle classed as each of 'cable PD', 'switchgear PD', 'unclassified', and 'noise' by existing knowledge rules.

Also used in some of the classifiers were details of every detected discharge above a minimum threshold of interest, including:

- position within the cycle (ms)
- signal peak (mV)
- magnitude of the discharge (pC)
- rise and fall times of the signal (ns)
- the width of the signal (ns)
- main frequency of the signal (MHz)

The process of summarizing the data was done by perl script, which populated new tables in the database keyed on the original channels for easy reference.

Data points in the months prior to a 'bad' point on the same channel were discarded so as not to pollute the 'good' pool, and completely silent channels were also discarded as being unhelpful in the construction of any type of classifier. Other data integrity checks are already performed as standard as part of IPEC's regular service.

## BAYESIAN CLASSIFICATION

The most significant approach attempted was to build a Naïve Bayesian classifier based on probability densities. Here, every input feature was assessed independently in RapidMiner with a temporary assumption of normal distribution, which builds equations to determine likelihood of failure based on that single feature. The resulting figures were then tested for the degree of overlap between the two classification groups ('good' or 'bad'), and those with sufficient predictive power were then used to re-build a Bayesian classifier by hand in Excel with parameters estimated by histogram in order to use the true distribution.

For the final predictions, these likelihoods were multiplied in combinations decided by analysis of a covariance matrix. Although Bayesian classifiers theoretically assume all parameters to be independent, they are often also highly

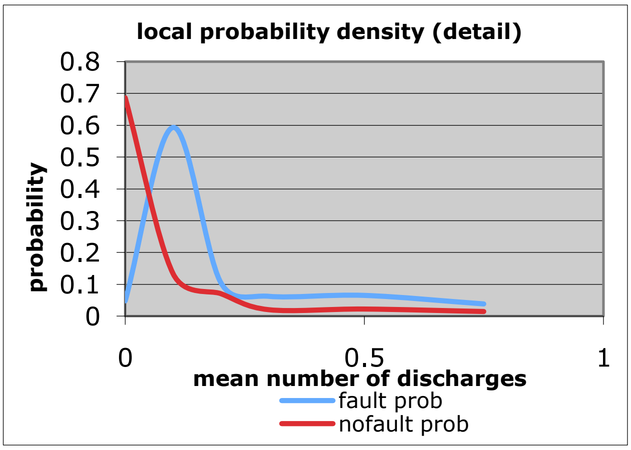


Fig 1. Probability density of local switchgear PD in channels with and without faults

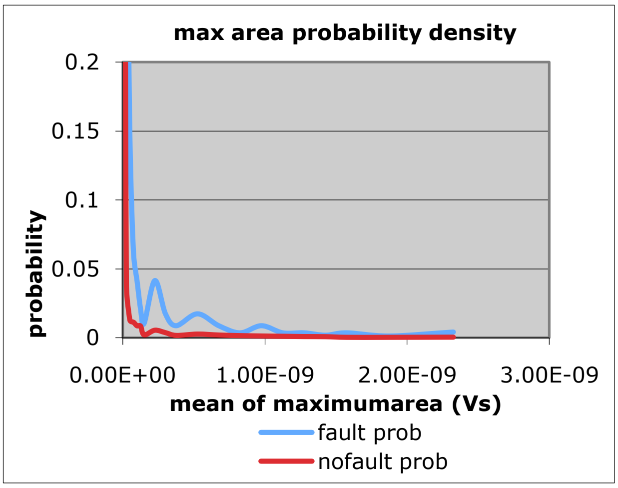


Fig 2. Probability density of cable discharge magnitude

accurate in cases where loose correlations exist between variables, and so the independence of the candidate features was balanced against their predictive powers.

## Testing

The resulting classifier was tested using threefold cross-validation - that is, the data was divided into thirds and each third was used in turn to test the accuracy of a classifier built using the other two thirds.

For each group, two different thresholds were used to determine what probability level should warrant being labelled. The first three used the global mean, and gave a decision for every tested data point. The second three (marked with an asterisk in Figure 3 below) had a higher threshold and refused to classify borderline cases, which gave it a higher accuracy.

| Group | Coverage | True +ive | False +ive | Lift |
|---|---|---|---|---|
| 1 | 100% | 42 | 149957 | 1.4 |
| 2 | 100% | 49 | 151814 | 1.6 |
| 3 | 100% | 46 | 141199 | 1.6 |
| 1* | 44% | 22 | 29977 | 3.6 |
| 2* | 61% | 28 | 33659 | 4.1 |
| 3* | 56% | 28 | 40113 | 3.5 |

Fig 3. Accuracy of Bayesian Classifier

'Lift' is given in the table as being the ratio of improvement over random chance in selecting positive samples, so a lift of 3.6 suggests a classfier will find 3.6 times as many failed channels as random. The large number of false positives is due to the extremely high ratio of 'good' to 'bad' samples. Lift in identification of negative samples is not shown, but was marginally above 1 in all cases.

We may also draw the conclusion from these graphs that **above a fairly moderate threshold of activity, there is little correlation between discharge levels and likelihood of failure.** A very strong, loud signal does not herald a significantly higher chance of failure than a medium signal, although both are considerably worse than a quiet channel. Whilst a small degree of this can be attributed to the early repair of the loudest channels, such channels were discarded rather than being flagged as 'good', and the effect is too pronounced for this to completely account for it.

## DYNAMIC TIME WARPING (DTW)

The difference between two time-varying sequences can be computed using a Dynamic Time Warping algorithm, which eliminates the effects of bad alignment, unmatched numbers of data points, or variations in the time scale of an interesting sequence. We used this method to produce a k-Nearest Neighbour classifier, in which an input to be tested was compared with a selection of known 'good' and 'bad' sequences. The percentage of 'bad' signals within the *k* closest sequences to the test signal was used as an indicator of its health.

A random sample was taken from the healthy pool in order to balance the issues with prior probability (healthy samples outnumber failed ones by a factor of several thousand). The classifier was assessed using leave-one-out validation – each data point was classified against the entire remaining data set, and the mean accuracy measured.

| Type | True +ive | False +ive | Lift |
|---|---|---|---|
| Max area 14 days | 19 | 36 | 3.1 |
| Total area 14 days | 8 | 28 | 2.0 |
| # cable PDs 14 days | 9 | 36 | 1.8 |
| # local PDs 3 days | 4 | 27 | 1.2 |
| Max peak 3 days | 21 | 109 | 1.4 |

Fig 4. Accuracy of DTW Classifier

Figure 4 shows the best performing input features. It was found that **cable PD trends to failure are more readily identifiable over 14-day periods, whereas switchgear failure is more sudden.** The lift is not as great as that of the Bayesian classifier, but is still non-trivial.

## DECISION TREES

Two data mining software packages, weka and RapidMiner, were used to construct decision trees based on the data using a variety of different algorithms. These included one-rule (1R), ID3 [6], and C4.5 (an improved ID3, implemented as 'J48' in weka).

It was found that a classifier built entirely on a single decision tree would be ineffective at best. However, the tree construction process was able to yield rules that had a very high success rate for positive predictions for a subset of the test group. That is to say, a very high percentage of true positives were made, at the expense of a large number of false negatives.

Of the rules found, the two most notable were

```
rise_maximumpeak > 0.001932 &&
avg_maximumpeak <= 0.000522: t (30.0)
```

from C4.5 (that is to say, **a large sudden rise in activity on a previously quiet channel is a predictor of failure**), for its combination of effectiveness and simplicity, and

```
avg_maximumpeak <= 0.000 &&
peak_maximumpeak > 0.004 &&
avg_nonlocal <= 1.388: true {t=29, f=0}
```

from ID3Numeric (low average activity both in recognized PD and unclassifiable borderline noise, with occasional spikes) for its extraordinary accuracy.

The first rule holds for 30 of the 168 total 'bad' samples, and represents an improvement ratio in picking true positives of 23.6 times over random chance when tested on the wider database. The second rule was even more impressive with a factor of 308. As with the first rule, this is limited by the large number of false negatives; in this case only 29 of the 168 fault samples were found. These rules may therefore usefully be used in conjunction with another classifier to provide an extra level of confidence on certain type of failure, but cannot be relied on alone.

Another thing we may conclude from the nodes of the constructed trees is that **switchgear discharge activity (the '…peak' and '…local' attributes) seem far more capable of predicting failure or otherwise than cable activity**.

## VISUALIZATION

Weka and RapidMiner both contain tools for plotting data across several axes, allowing one to change attributes and settings far faster and easier than the equivalent operations in Excel, gnuplot, or Matlab. Many combinations of input features were examined under various plot types, including XY scatter plots, coloured 3D scatter plots, density plots, and histogram surveys.
These were useful in identifying candidate features for classification (see earlier), and also produced a few unexpected results, the most interesting of which is shown here.

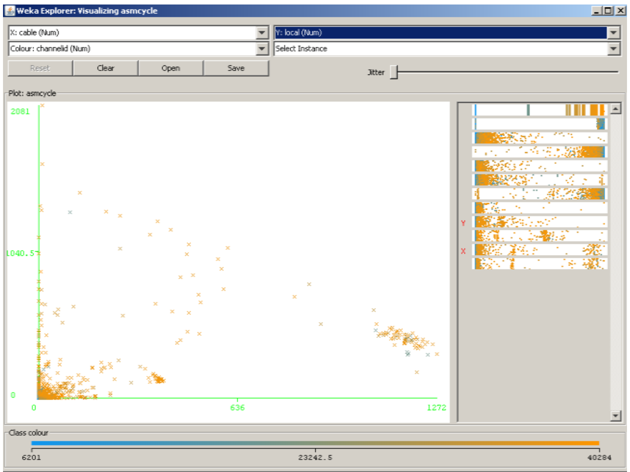


Fig 5. Screenshot of weka explorer illustrating cluster patterns

Here we see clustering in the relative levels of discharge detected as cable versus switchgear. Colour indicates channel id, i.e. the circuit being analysed – showing that the cluster on the right is a genuine pattern emerging in multiple channels, as opposed to a repeating pattern on the same channel.
The nature of this cluster is somewhat peculiar. The upper bound line running from top left to bottom right can be easily explained by the fact that the sum of types cannot exceed 2,000, therefore a high X value places a limit on the Y axis. However, the cluster appears in a very specific area even within this constraint, with no samples appearing at cable>1272, and only one below the local=400 mark. The smaller cluster at around 400x150 exhibits a similar ratio with a smaller magnitude.
We may conclude from this that in the data, very high quantities of cable PD are always accompanied by PDs recognized as switchgear PD. Whether this is a genuine physical property or an artefact of the sampling and recognition process is left as the subject of future investigation.

## OTHER METHODS

Association mining with algorithms such as *apriori* was attempted but was of little use, due to the fact that the results were flooded with the many obvious associations inherent in the data (high peaks in discharge imply high average pC values etc), obscuring any potential new insights or connections that may exist in the data set.

## ASSESSMENT AND OUTCOME

The classifiers were compared against one based on a simple rating of total picocoulombs of discharge, which broadly reflects the method currently used. The Bayesian classifier based on four-week-mean discharge was by far the most successful, with lifts of 4.1, 3.6, and 3.5 in a threefold cross-validation test. The DTW classifier also performed well, with a lift of 3.1 on 14-day patterns of discharge levels. Predictions of failure made by decision tree rules had lifts varying from 5 to over 300, albeit with low coverage.

Attempts were made to combine successful classifiers and input features, but these did not yield improvements in accuracy.

To summarize the main findings:

- From the probability densities: Magnitude and frequency of discharge occurrences are not proportional to likelihood of failure beyond a certain level, that is to say that a circuit with a very high level of discharge is not significantly more likely to fail than one with a medium level of discharge.
- From visualization: Unusual clusters can be seen in the frequency of discharge types within a circuit, even across channels from different substations. No explanation was found for their presence, however.
- From the decision trees: Many failures exhibit the property of having a very low average level of switchgear discharge but with sudden rises or occasional bursts of activity.
- From dynamic time warping: The behaviour of a circuit in the time leading to failure can potentially be recognizable from the pattern of its increases and decreases even when its absolute values are normalized. Also, failures in cable have a longer trend to failure than those in switchgear.

## REFERENCES

1. *Data mining techniques to assess the condition of high voltage electrical plant.* **McGrail, A.J., et al.** Paris : CIGRÉ, 2002. Vol. WG 15.11 of Study Committee 15.
2. **Babnik, T., Aggarwal, R. and Moore, P.** Data Mining on a Transformer Partial Discharge Data. *IEEE transactions on dielectrics and electrical insulation.* April 1997, Vol. 14, 2, pp. 444 - 452.
3. *Datamining for decision support of CBM of power cables.* **Gulski, E, et al.** Barcelona : CIRED, 2003. 17th International Conference on Electricity Distribution.
4. *Avoidance of MV switchgear failure case studies of on-line condition monitoring.* **Walton, Cliff, et al.** Prague : s.n., 2009. CIRED.
5. *Advances in continuous partial discharge monitoring.* **Smith, Colin, Fleming, Patrick et Michel, Matthieu.** Vienna : s.n., 2007. CIRED.
6. *Induction of Decision Trees.* **Quinlan, J.R.** 1, 1986, Machine Learning, Vol. 1, pp. 81-106.