Pilot Testing of an AI Algorithm to Identify Fault Category and Fault Cause from DFR Records

Finlay Macleod – System Performance Lead Engineer, Scottish Power
Mark Diamond – Product Manager, Qualitrol
Aaron Acton –Senior Data Scientist, Fortive Corporation
David Cole – Technical Application Specialist, Qualitrol
Paul Donnegan – Chief Architect, Qualitrol

UK and USA

Overview

Post event analysis of fault records to determine the distance to fault and the performance of protection and circuit breaker schemes has been an established method of working for many years. Experienced data Analysts, often protection engineers, normally undertake this task. Fault data must initially be downloaded from different sources, such as relays and stand-alone fault recorders, and converted to a common file format. Analysis usually consists of displaying the record in 'viewer' software and manually performing on screen measurements to pick out the salient features. Automating the process, where possible, to speed up the analysis has many advantages, faster access to accurate distance to fault allows the speedy dispatch of patrol teams, reducing the time to prepare a fault report allows Analysts to look at more records than they would have been able to in the past.

Scottish Power have a large installed base of fault recorders all reporting back to a central master station. Many fault records are generated daily especially during storm conditions. The effort required to analyse them is considerable despite the valuable insight they give to asset health. Not all records relate to circuit trips. This paper describes a pilot at Scottish Power to field test an AI algorithm to categorise 'fault type', for example trip, through fault, voltage dip, close onto fault etc as a precursor to 'standard' automation of measuring magnitude, time and calculating distance to fault. This allows the Analyst to filter, view a summary of the event and prioritise records to maximise productivity. Another feature of the algorithm is to determine fault cause. At present, two types are being tested, lightning and sub-cycle disturbances typical of VT problems. Other types are planned to be added.

The AI algorithm uses a Random Forest technique trained on a large data set of labelled records supplied by multiple Utilities.

Scottish Power have set up a separate master station accessing 23 x fault recorders in 13 substations. The devices were chosen as they are most likely to return the maximum number of fault records in a short time. Results so far have been more successful than previous rules based analysis. The trial will be extended to more substations in the future to gain more experience.

Scottish Power

The Scottish Power group of companies, owned by Ibedrola, operates in the UK. SP Transmission plc is a subsidiary of SP Energy Networks. It owns, operates and maintains infrastructure throughout Central and southern Scotland as shown in Fig 1.
SP Transmission takes electricity generated from power stations, windfarms and various other utilities and transports it through a large transmission network, consisting of over 3700 kilometres of overhead

lines and over 600 kilometres of underground cables. There are over 150 substations and in excess of 100 grid supply points in the network. The system is crucial to the delivery of the UK Government's renewable energy objectives due to its location in an area of outstanding renewable resource and its geographical location. SP Transmission has a unique role in connecting renewable generation and bulk transfer of renewable energy from Scotland into England & Wales. It operates assets at 400kV, 275kV, 132kV and 33kV at bulk supply points.

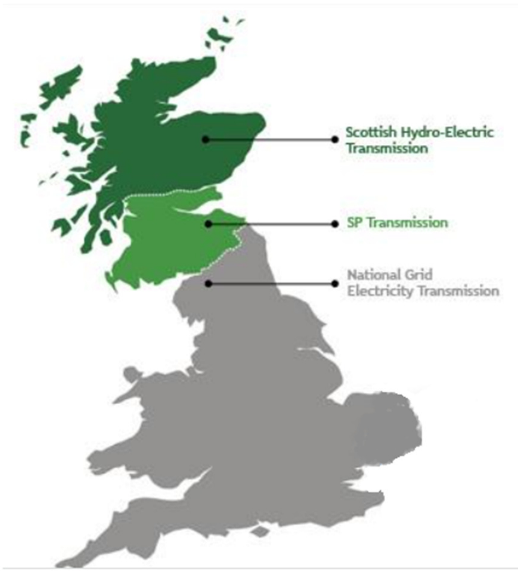

Fig. 1 SP Transmission Service Area

Scottish Power Monitoring Philosophy and Organisation

Scottish Power have been actively installing and operating a monitoring programme for over 25 years, A dedicated team is assigned to oversee the programme and analyse the data. The team are part of the PCM department that also includes the SCADA and protection engineers (Protection, Control and Monitoring).

To date, there are 687 Fault recorders installed across 202 sites. These return fault records, slow scan disturbance records, power quality and PMU data. In parallel, there is also a fleet of travelling wave fault locators providing accurate distance to fault results.
Findings from the subsequent analysis of the monitoring system are passed to the relevant departments, protection, asset maintenance or operations, for action. Findings from fault records can include protection relay / scheme issues, circuit breaker defects (slow operation, excessive pole spread), low frequency oscillations or advanced notice of VT problems.

All fault records are downloaded every night to a central server for archiving. In previous years, when there were just 60 devices, every fault record was proactively manually analysed by a team of two people. This is no longer possible. The number of recorders, hence the number or records, has grown and, despite having a larger team of six people, the role has expanded to include equipment to drive a condition based maintenance strategy. The team now look at fault records on a reactive basis, normally when the Control room notify when a trip or other incident has occurred.

The number of fault records generated depends largely on the prevailing conditions. In the extreme case of a storm there can be many hundreds generated over a few hours. Scottish Power trigger fault records from under voltage (90% of nominal), digital inputs and, at some 33kV sites, increased

residual current. Power swing triggers are set at major generating nodes, interconnectors, DC link connections etc. Rate of change triggering is also used at some sites to capture switching transients. A typical site during average conditions generated 60 fault records in 4 months. It is estimated that about 15% of these are 'useful'. The remainder will be a combination of commission tests, protection testing and 'far away' faults on the 400kV system triggering records over a wide area due to the resultant voltage dip.
If the numbers are extrapolated across the whole region it equates to 300 records per day of which 45 are useful.

The team still have the capability to analyse fault records after a trip but there is a worry about what is being missed by not taking the proactive approach on ALL records.

Benefits of Automatic Analysis of Fault Records

The Monitoring System returns information essential for the efficient operation of the power grid but there is too much to manually sort. The ideal is for an automated system that can return, for example:

- Information on protection performance
- Information on circuit breaker issues like pole spread, slow operation, I^2t, and trip coil analysis.
- Advanced warning of VT issues
- Assess voltage levels / power quality at rail / traction supply points and Grid Supply Points.

Fundamentals of an analysis system

A full record analysis system (RAS) featuring all the above is a big step to take in one release. As such, it is prudent to initially work on a reduced feature set and prove correct operation before adding extra functionality.

Earlier RAS software used a rules based method to categorise the type of fault record, for example, a trip, through fault, voltage dip, circuit switched out, circuit switched in, no fault and unknown. From that, the magnitudes of max and min voltage and current were measured and the protection and breaker operate times calculated from digital inputs. These values were published in a report along with a single end impedance distance to fault result. The object was to provide a first level of analysis to quickly find 'abnormal' behaviour and prioritise the work for manual analysis.

Trials with this approach were disappointing as the rules based method to categorise the fault was not reliable enough, it got it wrong too many times! An alternative method was needed.

Development of Machine Learning Algorithm

The objective was to develop an automated method for classifying fault records by fault category. A secondary objective was identifying root cause for circuit trips. Identifying root cause presented a greater challenge than record category partly because trips are relatively rare and acquiring enough confirmed examples for each of the different root causes is very difficult.

A large set of historical records, about 1.5 million, was collected from different transmission utilities and stored in Comtrade format. A small team of experts in fault record analysis examined a selection of the records and added labelling where necessary such that each record had a corresponding fault category. This annotated data, approximately 45,000 records, served as an initial dataset for use in a supervised machine learning system.

An exploratory data analysis was performed to begin to understand the dataset's characteristics, which presented several challenges for typical machine learning techniques.

Each Comtrade record could contain multiple channels from separate phases and circuits, have a unique duration, and be sampled at various rates. To address these factors, we converted the sampled waveforms to a common structure.

Using the labelled records and preprocessed signals, we trained a variety of machine learning models, reserving some portion as a validation set. Where model predictions differed from validation labels, experts were consulted to either support or correct earlier decisions. We found that a random forest model consistently performed with high accuracy and high performance.

Another challenge faced was the class imbalance caused by the occurrence rate of different fault types. Some faults were common, while others appeared only once in a thousand records or ten thousand records. To obtain additional examples of rare classes, we employed multiple iterations of active learning, in which our initial trained model generated predictions on unlabelled data, which then informed subsequent labelling rounds.

Our initial model was able to achieve a 90% accuracy across 10 fault categories and reached 97% after several refinement iterations with diminishing returns. Exceeding this performance level proved difficult, seemingly due to some inconsistency among labelers and ambiguity of fault categories, for example, at what point does a voltage dip become a through fault.

Some work was also undertaken on identifying two types of root cause where we managed to collect sufficient examples. These were lightning strike and VT issues. Any other unidentified root cause is classified as ‘unknown’.

Pilot Project and Interim Results

The pilot was set up in parallel with the SP main Master Station (MS) such that the experiment would not interfere with daily work. See figure 2.

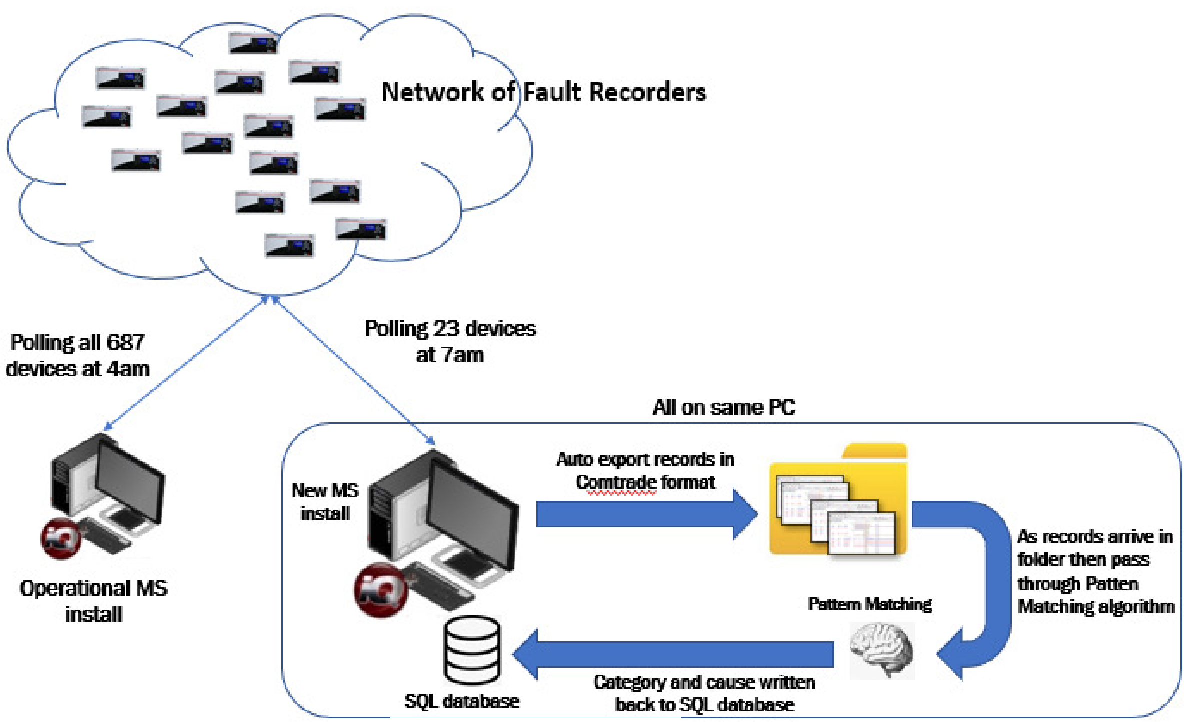

Fig. 2 SP Pilot Project

Twenty three devices from thirteen substations were selected on the basis there would be a high chance these would produce a greater number of varied fault records. The devices were polled daily to download records. A different time was chosen from the main master station to avoid any possible conflicts.

The machine learning (ML) algorithm was run outside of the master station to allow any updates from trial outcomes to be easily incorporated. The algorithm works from the comtrade file format to make it more generic to handle records from different sources. The master station auto exports records in comtrade format, the ML algorithm watches for new files and imports and processes them before returning the results back to the master station database where they can be included on the display of record details.

Figure 3 shows a screen shot of results from some historical records from a device's memory. The 'User Description' column has been temporarily used to show root cause. It is not surprising that all are classed as 'Unknown' as, to date, there are only two other alternatives, lightning strike or VT issue.
Even with this limited dataset it is easy to spot the 'less useful' records categorised as voltage dip, voltage swell and no fault. Filtering is possible on the 'Category' and 'User Description' columns meaning a user can select which categories and root causes are displayed at any time.

Data Analysis / Triggered Recordings

Substation Name	Device Name	Trigger Date & Time	User Description	Triggered Parameter Info	Classification	Time Quality	Fault Category
New Cumnock	BLAH_DUNH_2	13/12/2022 00:06:30.118	QAI cause: unknown	ROC ~ NECU BH DH2 Vr	FRSENSOR	Locked	voltage swell
New Cumnock	BLAH_DUNH_2	13/12/2022 00:02:38.419	QAI cause: unknown	ROC ~ NECU BH DH2 Vr	FRSENSOR	Locked	voltage swell
New Cumnock	BLAH_DUNH_2	13/12/2022 00:02:34.119	QAI cause: unknown	ROC ~ NECU BH DH2 Vr	FRSENSOR	Locked	voltage swell
New Cumnock	BLAH_DUNH_2	13/12/2022 00:02:32.478	QAI cause: unknown	ROC ~ NECU BH DH2 Vr	FRSENSOR	Locked	voltage dip
New Cumnock	BLAH_DUNH_2	13/12/2022 00:02:31.179	QAI cause: unknown	ROC ~ NECU BH DH2 Vr	FRSENSOR	Locked	voltage dip
New Cumnock	BLAH_DUNH_2	13/12/2022 00:02:29.459	QAI cause: unknown	ROC ~ NECU BH DH2 Vr	FRSENSOR	Locked	voltage swell
New Cumnock	BLAH_DUNH_2	13/12/2022 00:02:26.918	QAI cause: unknown	ROC ~ NECU BH DH2 Vr	FRSENSOR	Locked	voltage swell
New Cumnock	BLAH_DUNH_2	12/12/2022 15:20:41.569	QAI cause: unknown	ROC ~ NECU BH DH2 Vr	FRSENSOR	Locked	no fault
New Cumnock	BLAH_DUNH_2	12/12/2022 15:20:36.959	QAI cause: unknown	ROC ~ NECU BH DH2 Vr	FRSENSOR	Locked	no fault

Fig. 3 Example Display of Results

Two thousand one hundred records have been downloaded and analysed so far and work is in progress to manually check the accuracy of the ML algorithm. No errors have been detected to date which already demonstrates a superior performance to the previous rules based method.

It is encouraging that a VT issue has been correctly identified on new records captured since the start of the trial. This record showing an issue with the red phase can be seen in Fig 4. This would have been missed if the analysis was totally reliant on manual means.

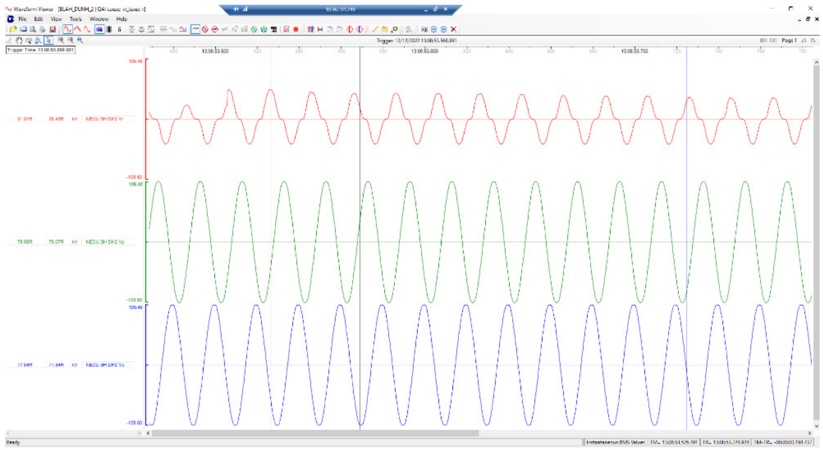

Fig. 4 Correctly identified VT Issue

Next Steps in the Medium Term

- Add more devices to the trial to obtain a better cross section of record types to further test the algorithm. Update the algorithm if necessary.
- Allow the fault category and root cause to be integrated into the RAS summary report in the master station software where voltage and current magnitudes, fault duration and timings of digital channels are listed. Such a summary will add benefit to the data analyst by better prioritising the records requiring more detailed study.
- Expand the list of identifiable root causes. The next type will most likely be tree or vegetation contact.

Summary

The success to date of the machine learning algorithm to correctly detect record category compared to previous rules based methods means progress can now be made on a record analyses package that better serves the needs of data analysts. The potential for identifying root cause is another important feature to determine the next steps in remedial actions that will contribute to reducing downtime and assist in the preparation of statistics for submission to the regulator.