**POE - New Enhancement Work 1**

July 2022

# **Summary**

Per discussion with NR SMEs on the requirement of enhancing the current POE “Defect Prediction Model”, various approaches were considered

The approaches were:

* Considering other routes including LNW
* Building new features based on current data
* Keep some existing features and tweak (if required)
* Building model considering different types of algorithms (apart from existing algorithm which was used in current POE Model)
* Test the model on different time periods of dataset

Data Science team was progressing on building new features, tweaking existing features on different routes. However, POE New Enhancement Work activity has been paused due to other priority tasks (to the Data Scientist who were working on POE Model) from NR SMEs.

The below sections showcase the work done in POE New Enhancement Work.

# **In Scope - Data**

* Asset Type considered: Clamp Lock Mark 2
* Data Period: Jan 2017 to June 2020
* Routes Considered: LNW, Kent, Mid-East & Anglia

# **Feature building**

In POE New enhancement Work, initially per discussion with NR SMEs, total of 12 features were finalised, out of those 3 features were from Phase 1 of the traces, 4 features were from Phase 2 of the traces, 1 feature was from Phase 3 of the traces and finally 4 features not based on the phases of the traces. However, out of 12 features, 7 features were new, and 5 features were either existing (from the previous POE Model) or required some tweaking on the existing features.

Table 1 shows the features and the description of the individual features mentioned below:

**Table 1. Features and their description**

|  |  |  |  |
| --- | --- | --- | --- |
| **Trace Phase** | **New Feature Name** | **Feature Description** | **Feature Category (New/Existing/Tweaking)** |
| **Phase 1** | Ph\_1\_Max\_Current\_Deviation | Deviation of maximum current of a trace in Phase 1 with the average maximum current of the particular asset of Phase 1 | New |
| Ph\_1\_Median\_Current\_Position | Capture the position of the swing time where the median line touches the current value (Y-axis) | New |
| Ph\_1\_Min\_Current\_Deviation | Minimum current deviation in first phase | Existing |
| **Phase 2** | Ph\_2\_Max\_Current\_Location | Capture the timing of the swing time where the Maximum current touches the Y-axis (current value) | New |
| Ph\_2\_Max\_Current\_Location\_Deviation | Capture the deviation of timing of the swing time of the trace where the Maximum current touches the Y-axis (current value) with that of Average Timing of the swing time of the asset | New |
| Ph\_2\_Max\_Current\_Deviation | Deviation of maximum current of a trace in Phase 2 with the average maximum current of the particular asset of Phase 2 | Tweaking |
| Ph\_2\_Average\_Current\_Deviation | Difference of maximum current in the Phase 2 of the trace with average current of Phase 2 of the asset | Existing |
| **Phase 3** | Ph\_3\_Max\_Current\_Deviation | Deviation of maximum current of a trace in Phase 3 with the average maximum current of the particular asset of Phase 3 | New |
| **Others** | Swingtime\_Deviation | Deviation of the swing time of the trace with the average swing time of the particular asset | Tweaking |
| Swingtime\_Difference\_From\_Previous\_Trace | Swing time difference from the previous trace | Existing |
| Weather\_Condition | Girish Sabbi is working on it | New |
| Pattern\_Matching\_DTW | Trace pattern matching Using DTW | New |

# **Visualisation of features**

Few features have been selected from different phases of a trace for visualisation purpose.

## **4.1 Features from Phase 1 of Trace**

**i) Ph\_1\_Max\_Current\_Deviation:** Deviation of maximum current of a trace in Phase 1 with the average maximum current of the particular asset of Phase 1

Graphical user interface

Description automatically generated

**Figure 1. Visualisation of feature - Ph\_1\_Max Current Deviation**

**ii) Ph\_1\_Median\_Current\_Deviation:** Capture the position of the swing time where the median line touches the current value (Y-axis)

Graphical user interface

Description automatically generated

**Figure 2. Visualisation of feature - Ph\_1\_Median Current Deviation**

## **4.2 Features from Phase 2 of Trace**

**iii) Ph\_2\_Max\_Current\_Location:** Capture the timing of the swing time where the Maximum current touches the Y-axis (current value)

Graphical user interface

Description automatically generated

**Figure 3. Visualisation of feature - Ph\_2\_Max Current Location**

**iv) Ph\_2\_Max\_Current\_Location\_Deviation:** Capture the deviation of timing of the swing time of the trace where the Maximum current touches the Y-axis (current value) with that of Average Timing of the swing time of the asset where the Maximum current touches the Y-axis (current value)

Graphical user interface

Description automatically generated with medium confidence

**Figure 4. Visualisation of feature - Ph\_2\_Max Current Location Deviation**

**v) Ph\_2\_Max\_Current\_Deviation:** Deviation of maximum current of a trace in Phase 2 with the average maximum current of the particular asset of Phase 2

Graphical user interface, text, application

Description automatically generated

**Figure 5. Visualisation of feature - Ph\_2\_Max Current Deviation**

## **4.3 Features from Phase 3 of Trace**

**vi) Ph\_3\_Max\_Current\_Deviation:** Deviation of maximum current of a trace in Phase 3 with the average maximum current of the particular asset of Phase 3

Graphical user interface

Description automatically generated with medium confidence

**Figure 6. Visualisation of feature - Ph\_3\_Max Current Deviation**

**vii) Swingtime\_Deviation:** Deviation of the swing time of the trace with the average swing time of the particular asset

Chart

Description automatically generated with medium confidence

**Figure 7. Visualisation of feature - Swingtime Deviation**

# **Data Quality Issues (DQI)**

On exploring the features, some data quality issues (in addition to the previously observed data quality issues) were observed in the dataset. Below are few issues mentioned:

## **5.1 DQI - Traces having same timestamp but different trace direction**

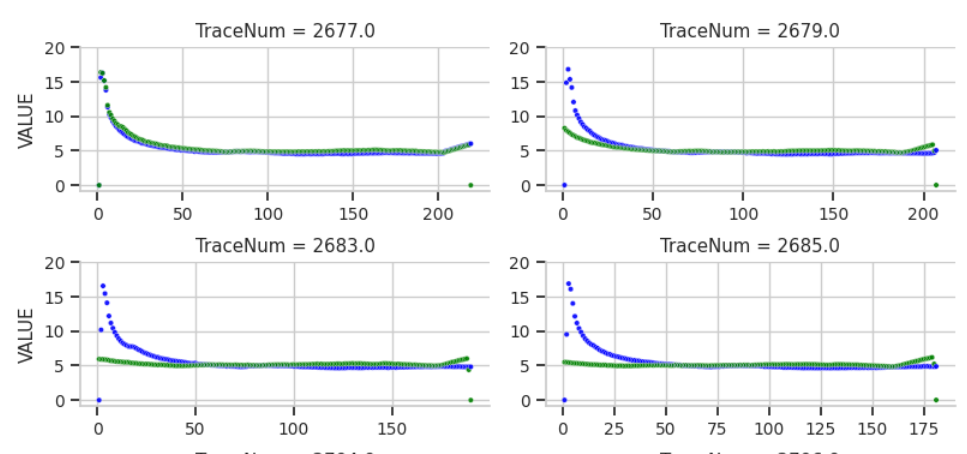
Table 2 shows the data quality issue were traces with same timestamp but different trace direction.

**Table 2 Traces having same timestamp but different trace direction**

**Table

Description automatically generated**

Figure 8 Shows the visualised example of the traces with same timestamp but different trace direction.

****

**Figure 8. Example of traces having same timestamp but different trace direction**

## **5.2 DQI – Current range changes after failure**

In some assets, it was observed that the current range changes after failure which reflected in some of the features as a result it will be difficult to build consistent features for these types of assets.

Figure 9 & 10 shows example of current range changes after failure for Ph\_3\_Max\_Current\_Deviation and Swingtime\_Deviation feature.

**Chart

Description automatically generated**

**Figure 9. Example of current range changes after failure for Ph\_3\_Max\_Current \_Deviation feature**

Chart, scatter chart

Description automatically generated

**Figure 10. Example of current range changes after failure for Swingtime \_Deviation feature**

# **Other Observation – Seasonality**

While creating and visualising the features, some of the features show some cyclical pattern. Figure 11, which is Phase\_3\_Max\_Current\_Deviation shows the evidence that there is upward trend from the month of September to January every year and downward trend from February to August every year.

Chart, scatter chart

Description automatically generated

**Figure 11. Evidence of seasonality in some features**

However, similar type of cyclical pattern has been observed in many others features like Phase\_2\_Max\_Current\_Deviation, Swingtime\_Deviation.

However, the issue can be solved by calculating the averages for the features quarterly instead on average on full period data

# **Algorithms for measuring the similarity between two temporal sequences of traces**

## **7.1 Objective**

The objective of this exercise was to identify which phase (Unlocking/Transition/locking) phase of the trace was showing more dissimilarity with the ideal trace and hence would help us determine the most problematic section of the experimental trace.

**Ideal Trace:** Ideal trace was obtained for each asset and for each direction (RN/NR) with the help of the isolation forest algorithm.

**Experimental Traces:** All the traces apart from the ideal traces are referred as the experimental traces.

Chart, line chart

Description automatically generated

**Figure 12. represents an ideal trace.**

## **7.2 The approach:**

Three approaches were initially considered to carry out this exercise. The approaches have discussed in subsequent stages.

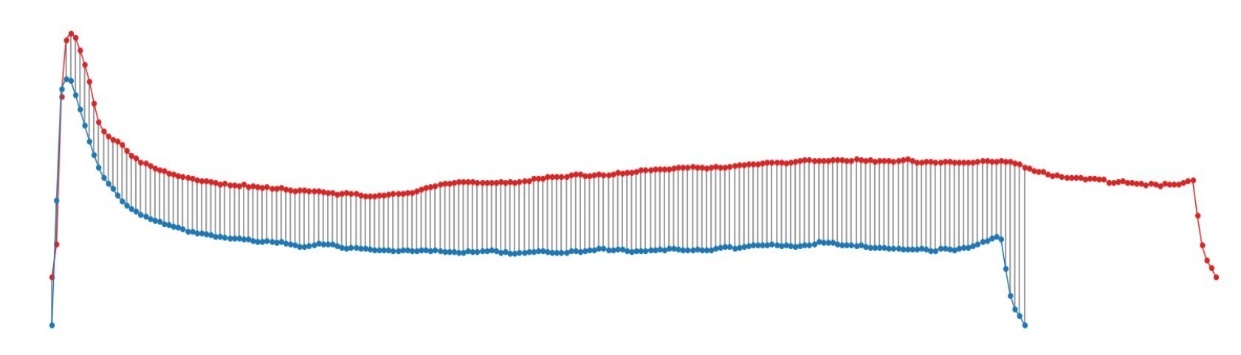
### **7.2.1 Approach 1: Comparison using the Euclidean Distance**

Here we used the Euclidean distance algorithm to compare the individual points of the reference trace and the ideal trace.

***Challenges:***

* Trace comparisons using the Euclidean distance where the number of data points in the experimental trace and the ideal trace were not the same as shown in figure 13, as Euclidean distance follows a one-to-one mapping approach to calculate the distance from the individual points from both the traces.
* This would eventually leave out a few data points from either trace to be compared.

Figure 13 shows the above approach would leave out few data points for comparison if the trace length is dissimilar



***Ideal Trace***

***Experimental Trace***

**Figure 13. shows the comparison using Euclidean distance algorithm**

### **7.2.2 Approach 2: Complete alignment of experimental trace with the ideal trace using DTW Algorithm.**

* In this approach we used the DTW algorithm to completely align the experimental trace and the ideal trace.
* This helped us in obtaining the alignment cost of both the traces which was a numerical representation of the dissimilarity between the two traces.

***Challenges:***

* The mapping exercise helped us obtain the alignment cost of the experimental trace with the ideal trace, however the mapping was not consistent.
* Figure 15 shows that the mapping was done in such a manner that a data point in the second phase of the experimental trace was mapped to the point in the third phase of the ideal trace.
* This was inconsistent with the objective as it wouldn’t guide us to the phase of the trace which was posing more problems.

Diagram

Description automatically generated

**Figure 14: Shows that in the ideal scenario where the ideal trace and the experimental trace are identical the alignment cost would be zero.**

Chart, line chart

Description automatically generated

**Figure 15. shows the non-coherent trace comparison where a point from the second phase is being mapped to third phase**

### **7.2.3 Approach 3: Phase wise alignment of the experimental trace with the ideal trace.**

* In this approach we segmented the experimental trace and the ideal trace beforehand.
* We then applied the DTW algorithm on the individual phases of the experimental trace and the ideal trace and obtained the alignment cost for each phase of the traces.

***Benefits :***

* The above approach eliminated the problem where a data point from a different part of the experimental trace was mapped to a different phase of the ideal trace and vice-versa

# **Overall Observation**

* Ph1\_Max\_Current\_Deviation is showing very less deviation before failure
* Ph2\_Max\_Current\_Deviation, Swingtime\_Deviation and Ph3\_Max\_Current \_Deviation are showing some deviation before failure
* To calculate the Ph1\_Median\_Current\_Position, first 4 datapoints have been ignored to overcome the challenges of calculating the median current.
* Ph1\_Median\_Current\_Position is showing very little correlation with failure based on feature visualiation
* Ph2\_Max\_Current\_Location and Ph2\_Max\_Current\_Location\_Deviation are showing negligible correlation with failure based on feature visualisation.