



Wang, Xiaoyu and McArthur, Stephen D.J. and Strachan, Scott M. and Kirkwood, John D. and Paisley, Bruce (2017) A data analytic approach to automatic fault diagnosis and prognosis for distribution automation. IEEE Transactions on Smart Grid. ISSN 1949-3053 , <http://dx.doi.org/10.1109/TSG.2017.2707107>

This version is available at <https://strathprints.strath.ac.uk/60710/>

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Unless otherwise explicitly stated on the manuscript, Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Please check the manuscript for details of any other licences that may have been applied. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (<https://strathprints.strath.ac.uk/>) and the content of this paper for research or private study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to the Strathprints administrator: strathprints@strath.ac.uk

A Data Analytic Approach to Automatic Fault Diagnosis and Prognosis for Distribution Automation

Xiaoyu Wang, *Student Member, IEEE*, Stephen D.J. McArthur, *Fellow, IEEE*, Scott M. Strachan, *Member, IEEE*, John D. Kirkwood and Bruce Paisley

Abstract—Distribution Automation (DA) is deployed to reduce outages and to rapidly reconnect customers following network faults. Recent developments in DA equipment have enabled the logging of load and fault event data, referred to as ‘pick-up activity’. This pick-up activity provides a picture of the underlying circuit activity occurring between successive DA operations over a period of time and has the potential to be accessed remotely for off-line or on-line analysis. The application of data analytics and automated analysis of this data supports reactive fault management and post fault investigation into anomalous network behavior. It also supports predictive capabilities that identify when potential network faults are evolving and offers the opportunity to take action in advance in order to mitigate any outages.

This paper details the design of a novel decision support system to achieve fault diagnosis and prognosis for DA schemes. It combines detailed data from a specific DA device with rule-based, data mining and clustering techniques to deliver the diagnostic and prognostic functions. These are applied to 11kV distribution network data captured from Pole Mounted Auto-Reclosers (PMARs) as provided by a leading UK network operator. This novel automated analysis system diagnoses the nature of a circuit’s previous fault activity, identifies underlying anomalous circuit activity, and highlights indications of problematic events gradually evolving into a full scale circuit fault. The novel contributions include the tackling of ‘semi-permanent faults’ and the re-usable methodology and approach for applying data analytics to any DA device data sets in order to provide diagnostic decisions and mitigate potential fault scenarios.

Index Terms—Distribution automation, distribution network data, decision support system, fault activity, fault diagnosis and prognosis, pick-up activity.

I. INTRODUCTION

AS load grows and regulatory regimes focus increasingly on the security of supply and reliability [1], distribution systems have become much more complex to plan, control, and

maintain. Therefore, in order to effectively manage these aspects of network operation, major investments and developments have been undertaken in the area of Distribution Automation (DA). In particular, DA is applied for the purpose of automatic sectionalising and localisation of network faults and improving restoration times, with an aim to reduce Customer Minutes Lost (CML) and Customer Interruptions (CIs) [2], and generally improve customer service.

The ongoing implementation of DA has led to increasing volumes of operational data becoming available from Intelligent Electronic Devices (IEDs) [3] and other DA devices. These data sets can be analysed to provide fault diagnosis and network event information and predictions, but are not always capitalised upon. There have been attempts to use the data through diagnostic algorithms within automated decision support systems [4].

Previous research has delivered diagnostic capability in related areas. The data captured from Supervisory Control and Data Acquisition (SCADA) or Advanced Metering Infrastructure (AMI) combined with DA devices has been automatically analysed using intelligent system methods [5]. This has provided fault analysis and diagnostic assistance for protection engineers [6][7]. There has also been a focus on fully automating diagnostics and integrating these into operational systems [8][9], and the use of novel approaches for network monitoring [10].

Meanwhile, other research has concentrated on the practical use of IED data as the primary source to support automated data analysis to assist engineers. This has focused on monitoring system behavior and improving system operation [11][12]. Popovic et al. [13] proposed a fully automated data analytics solution to improve decision-making processes by interpreting the network data retrieved from digital protective relays.

This research takes the next step. It automates the analysis of Pole Mounted Auto-Recloser (PMAR) data to diagnose PMAR faults. In addition, it automatically identifies emerging circuit and device faults to allow preventative measures to be taken to avoid or minimise outages. Specifically, the contributions of this paper are: research, design and proof of a predictive capability to identify emerging faults within a DA application; the combination of data science and knowledge based techniques to deliver diagnostic and prognostic functionality; a unique focus on distributed DA autoreclosing devices which

This work was supported by SP Energy Networks.

X. Wang, S. D. J. McArthur, and S. M. Strachan are with the Institute for Energy and Environment, Department of Electronic and Electrical Engineering, University of Strathclyde, UK. (e-mail: xiaoyu.wang@strath.ac.uk, s.mcarthur@strath.ac.uk, scott.strachan@strath.ac.uk).

J. D. Kirkwood and B. Paisley are with Engineering Services of SP Energy Networks, UK. (e-mail: bruce.paisley@spenergynetworks.co.uk).

allows the identification of evolving or incipient network faults; and, the full implementation of a prototype and evaluation against actual industry data. This research addresses a gap in the field of power system data analysis, which is the use of Low-Voltage Distribution Network IED data for fault diagnosis and prognosis. The methods can be adapted for other DA devices.

II. CIRCUIT PERFORMANCE AND UNSOLICITED OPENINGS

A PMAR placed on the overhead lines of a distribution network [14] is an example of a protection IED [15] which combines monitoring and communication capabilities. It automatically isolates faulted sections of an overhead line circuit, while simultaneously capturing the fault details and storing them in log files (e.g. current and voltage amplitude, affected phase, etc.) [16]. Fig. 1 diagrammatically shows the operation of the device.

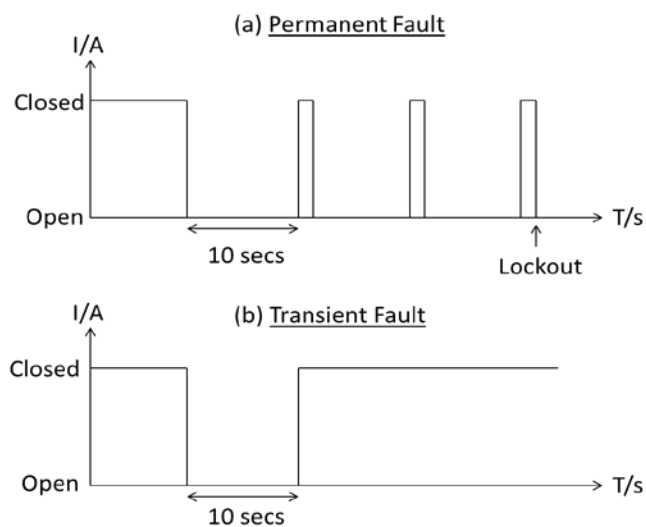


Fig. 1. PMAR operation

When a fault occurs, the PMAR attempts a set number of reclosure attempts (typically 3 times, as set by the operator in this study). If a reclosure attempt is unsuccessful (indicating the continued presence of a fault) the PMAR device remains open for a period of 10 seconds before attempting a reclose. If the fault remains on the circuit during all three reclosure attempts, the PMAR will ‘lockout’ on its third attempt, suggesting that a ‘permanent fault’ (or non-transient fault) may exist on the circuit (see Fig. 1(a) above). Following a ‘lock-out’, the PMAR can only be reclosed manually via telecontrol by the control engineers. Therefore, it is clear that a PMAR can only remain closed if the fault has been removed from the circuit or isolated from the PMAR in question. Fig. 1(b) shows the operation of the PMAR after a short-lived ‘transient fault’ has appeared and then almost immediately been removed from the circuit. Here the PMAR trips when the fault is initially detected and, as the fault clears during the 10 second period of isolation, the PMAR remains closed after its first attempt at reclosure.

‘Semi-permanent faults’ are intermittent faults that go undiagnosed [17]. An example cause of such a fault is rain affecting a cracked insulator on a wood pole (which can dry out

and then no longer provide a fault path). These can adversely affect the quality of daily electricity service through ‘nuisance tripping’, which incur regulatory penalties for network operators. Semi-permanent faults result in unsolicited PMAR trip activity, referred to as ‘Unsolicited Openings’ (UO) by the network operator.

From a technical/asset management perspective, the frequency of UOs on a particular circuit could provide an indication of deterioration in a circuit’s (or indeed an individual PMAR’s) performance and underlying condition. In addition, the corresponding log files generated by the PMAR device may contain ‘pick-up’ activity (line current and voltages) providing a deeper insight into the device/circuit condition and evidence of evolving fault conditions. For example, the cracked insulator mentioned previously could ultimately cause an initial PMAR trip when moisture ingresses the crack. As the PMAR is then reclosed (in some instances multiple times) this can cause the moisture to evaporate and the PMAR to remain closed after a reclosure attempt. This restores the supply to the circuit for a time until the scenario repeats itself in the future. As the circuit supply is restored, there is usually the assumption that some prolonged transient fault has cleared with no lasting damage to the circuit, but this may not in fact be the case. In addition, from a customer service perspective, assessment of a circuit’s UO activity provides an indication of which customers may be most prone to ‘nuisance tripping’, compromising the level of customer service they experience. This exposes the Distribution Network Operator (DNO) to regulatory penalties.

The PMAR device generates log files containing overhead line current and voltage data sampled at 12.8 kHz. This is referred to as ‘pick-up’ data. It has a Main Processor Module (MPM), and along with its core protection functionality, it also records the sampled data and generates alarms based on its internal operation.

For the purpose of diagnosing this root cause and predicting evolving faults, this research focuses on the design and implementation of a Decision Support System (DSS) that assists engineers in recognising PMAR behaviour. This is based on analysis of ‘pick-up’ data stored in PMAR log files. The DSS has been developed and deployed as a Knowledge-Based System (KBS) [18][19] with the objective of diagnosing PMAR device faults. It also provides detection and early warning of semi-permanent circuit faults responsible for UOs, which threaten to evolve into more serious permanent faults. These adversely impact on customer service, and the data analysis functions to combat this are detailed in Section III. The DSS includes visualisation tools. These support the analysis of circuit behavior by engineers and the development of predictive and diagnostic rules, as shown in the Case Studies in Section V.

III. DATA ANALYSIS FOR DIAGNOSTIC AND CLASSIFICATION FUNCTIONS

This section outlines how the KBS generates and implements diagnostic rules from expert knowledge and engineers’ experience. It also describes how it uses the data analysis methods and approaches to diagnose PMAR device faults and

detect semi-permanent faults on the circuits. This functionality, in conjunction with the predictive capability for fault prognosis described in Section IV, was built into the prototype software DSS described in Section V.

A. Diagnosis of PMAR Device Faults

The condition monitoring of the PMAR is supported by its integrated Main Processor Module, which also controls the operation of the PMAR. The MPM records details of on-line activity (i.e. fault current and voltage amplitudes, etc.) and the condition of the PMAR's components. These recorded conditions or details contain the information on existing/potential device faults or circuit faults.

Therefore, after the raw log file data is imported, the knowledge-based system (KBS) automatically invokes the diagnostic rules (based on the experience of operational engineers), which characterise known PMAR faults. The aim is to prevent potentially delayed or failed operations in response to overhead line faults. These rules operate on the PMAR log file data [20] and are shown in Fig. 2.

Rules for diagnosing PMAR device faults	
#1	MPM fault diagnosis If Number of detected consecutive "MPM fault" alarms \geq 20 Then MPM fault
#2	Driver module fault diagnosis If "Excessive To" alarms Or "Excessive Tc" alarms are detected Then Driver Module fault
#3	PMAR tank fault diagnosis If Number of consecutive driver status changes (Open to Close time $<$ 10s) with "Undefined" alarms \geq 10 Then PMAR tank fault
#4	Microswitch fault diagnosis If Time interval between Action Trip and Status Open $<$ 25ms And Time interval between Action Close and Status Closed $<$ 40ms Then Microswitch fault

Fig. 2. Rules for diagnosing PMAR device fault

As illustrated in Fig. 2 there are four main device faults which can be diagnosed and reported to engineers. These faults are identified through: the interpretation of alarms generated by the MPM; or, calculations on interval times between the PMAR operations and status changes within the log file. As an explanation:

- An MPM fault could be confirmed by detecting alarms in the log file, which were generated automatically by the MPM.
- A typical driver module fault always leads to the failure of the PMAR operation for a fault within the protected zone. The KBS identifies this when the 'Excessive To' (i.e. contact opening time exceeds setting time) or 'Excessive Tc' (i.e. contact closing time exceeds setting time) events are indicated in the log file.
- A tank fault is characterised by multiple open and close status changes within impracticable time periods (i.e. duration from Open to Close is less than 10 seconds). The MPM recognises incorrect driver operation and

generates the alarm "Undefined".

- A microswitch fault generally reflects a problem in the PMAR's switch. The KBS identifies this fault by comparing the actual time periods of status changes with primary settings through automatic time stamp calculations.

This is the first stage of automated decision support, which focuses on PMAR device fault diagnostics. The next stage of the system concentrates on overhead line fault diagnoses.

B. Semi-Permanent Fault Detection on Circuits

As discussed previously, a Semi-Permanent Fault (SPF) manifests itself through sporadic periods of intense PMAR operation on a circuit. It is therefore necessary to first classify PMAR log file data into different classifications of PMAR operation. The next step is to filter the activity consistent with SPFs. Trends of classified SPF candidates can then be used to confirm the existence of SPFs. To assist in the detection of potential SPFs, a further rule-base was designed to automatically classify the PMAR operations being experienced. The rules were developed through engineers' expert knowledge and knowledge of the PMAR's operating mechanism. Following classification, a visualisation tool is used to allow engineers to observe the results and allows them to identify data trends and patterns associated with PMAR operations that could be due to potential SPF activities. In addition, data can also be visualised to substantiate the diagnosis of invoked rules. The following sections describe how the rules are generated and the trends evaluated to detect SPFs. The visualisation tools are described in Section V.

a) Classification of PMAR Operations

Using the log file 'pick-up' data recorded during the operation of PMARs under different fault conditions, four specific rules were defined. These were based on the experts' knowledge and understanding of how these faults are likely to manifest themselves in these data sets. The rules are then implemented within the KBS to group 'pick-up' activity into different classifications of PMAR operation depending on the pick-up duration and the number of corresponding PMAR operations. These are *fault 'pick-up'* activity resulting in: no trip (FP); *single trip* (ST); *multiple trip* (MT); and, *lockout* (L). Fig. 3 shows the rules for classifying PMAR operation (i.e. FP, ST, MT and L) in order of increasing severity.

Using these rules the DSS can then identify, map out and prioritise potential SPF activity on circuits which have experienced pick-up activity exceeding acceptable time limits. From this prioritised mapping of potential SPF activity across the network, experts can then drill into the data using the visualisation tool for a deeper analysis of the pick-up data. This allows them to discern the details of PMAR operation against phases affected, evidence of earth faults, etc. The Case Studies in Section V show trends of interest identified using the visualisations.

Rules for classification of PMAR operation	
#1	Abnormal 'pick-up' activity without leading to PMAR operation
If	The 'pick-up' activity's duration time > 30ms
And	disappear without PMAR's operation
Then	The classification is determined as FP
#2	Abnormal 'pick-up' activity leading to a single trip
If	The 'pick-up' activity's duration time > 30ms
And	cleared by a single PMAR's operation
Then	The classification is determined as ST
#3	Abnormal 'pick-up' activity leading to multiple trips
If	The 'pick-up' activity's duration time > 30ms
And	cleared by multiple consecutive PMAR's operations
And	without leading to a lockout
Then	The classification is determined as MT
#4	Abnormal 'pick-up' activity leading to a lockout
If	The 'pick-up' activity's duration time > 30ms
And	isolated by a PMAR lockout
Then	The classification is determined as L

Fig. 3. Rules for classification of PMAR operation

b) Evaluation of Behavioural Trends for SPF Detection

Following the classification of PMAR operation, their trends are then analysed. This provides an insight into potential SPFs. Moreover, the statistical evaluation of 'pick-up' activities associated with these PMAR operations could substantiate the existence of potential SPFs. The trends could directly reflect the evolving conditions associated with this form of fault. Fig. 4 shows an example of the trends and statistical features for considering FP PMAR operation (as detailed in Fig. 3). These support engineers in determining whether a SPF exists on the circuit. These trends and statistical features are applied to other classifications of PMAR operation (i.e. ST, MT and L) as well.

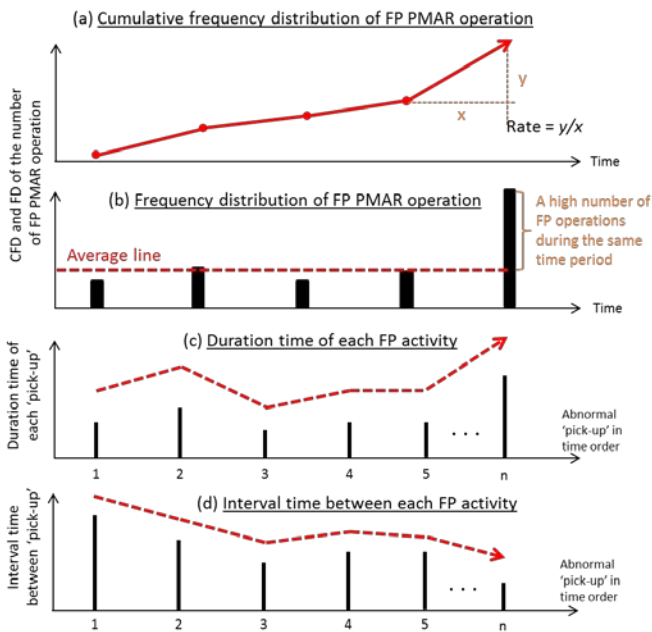


Fig. 4. Trends of statistical features

With respect to Fig. 4, the DSS will calculate and demonstrate the frequency of a number of FP operations for a specific PMAR. Where there is an increasing rate in the cumulative frequency distribution (CFD) (as shown in Fig. 4(a)), this suggests the FP operations resulted from faults

becoming more frequent and the corresponding frequency distribution (FD) (as shown in Fig. 4(b)) exceeds the average number of PMAR operations. The high number of FP operations during the same time period also indicates an evolving fault on the circuit (i.e. a SPF). To confirm the existence of the SPF, the system extracts the duration time (DT) of each associated 'pick-up' activity and determines the interval time (IT) between consecutive pick-ups through calculation. Trends showing an increase in DT and a decrease in IT suggest the circuit's condition is worsening. This may also be indicative of the stage of maturity of the semi-permanent fault in terms of its closeness to becoming a permanent circuit fault. Using the visualisation tool, the users obtain a perspective on the evolution of faults and can make informed and faster decisions regarding the existence of SPFs within the network.

IV. PROGNOSTIC FUNCTIONS

The KBS described in the previous sections can assist control engineers by diagnosing PMAR device faults and identifying potential SPFs present on overhead line circuits. In addition to this functionality, further work has been conducted into developing a prognostic capability which can predict potential PMAR operations (or evolution of SPFs into permanent faults). It grades the circuit's pick-up activity in terms of the imminence of such a threat. This would enable maintenance staff to take evasive action and potentially avoid expensive and prolonged outages required to repair damage from permanent faults.

In developing this prognostic capability, this paper proposes a structured method of data mining [21] to derive rules operating on anomalous pick-up activity. This is defined as a pick-up duration of greater than 30ms which has not yet led to a PMAR operation (previously classified in Fig. 3 as "FP"). These rules will then be capable of predicting future PMAR operations (i.e. classified as ST, MTs, and L in Fig. 3). The following sections detail the process of data mining applied to the analysis of FP pick-up activity data to generate the predictive rules. This can be reapplied to other DA and operational devices.

A. Data Preparation

FP activity is grouped into one-month time windows, which do not contain any recorded trip or lockout activity. The short one-month window of interest was selected to focus in on activity resulting from the same underlying cause (semi-permanent fault condition).

The training data set was produced from 12 PMAR log files. This represents pick-up activity captured at a sampling rate of 12.8 kHz across 7 circuits, spanning a period of up to 5 years. A total of 100 FP groups were derived from this training data set and subjected to the data mining process discussed in Section IV.B below.

The features selected for the predictive algorithm were:

1. Number of Total Pick-ups (NTP) in each FP group.
2. Time to Trip (TTT), which is the time duration from the last recorded pick-up activity in the FP group to the next PMAR operation.

3. Interval Time Trend (ITT) is a Boolean value which represents the increasing or decreasing of time interval between two consecutive fault pick-ups in a group of activity (i.e. True for decreasing, False for not decreasing).
4. Average Current Amplitude (ACA), which is the average current amplitude of FPs in each group.
5. Average Pick-up Duration (APD), which is the average pick-up duration of FPs in each group.

B. Data Segmentation and Visualisation

Since the PMAR's operation associated with the five defined features is unlabelled, data-class associations are unknown. In order to extract hidden associations between the data characteristics and those of FP activities, clustering techniques were applied.

The K-Means algorithm is a clustering technique used for data mining which is simple to implement and straightforward to interpret. It partitions similar unlabelled data (i.e. data without defined categories or groups) into a pre-set number (K) of clusters [22]. When compared with alternatives, such as the hierarchical clustering algorithm [23], K-Means provides increased accuracy. That is, it has more possibilities of partitioning the data point to the correct allocated cluster when the relative weights of the features are not well understood. Therefore, the application of a K-Means algorithm enables segmentation of the data into distinct clusters where specific clusters could be considered indicative of distinct PMAR operating conditions [24].

To visualise the clustering output from the K-Means algorithm, a dimensionality reduction technique is required to process the clustered data. The t-Distributed Stochastic Neighbour Embedding (t-SNE) methodology is one of a set of algorithms that can transform high-dimensional data into two or three dimensions [25]. Other techniques, such as Sammon Mapping [26], can be used for this purpose. The t-SNE algorithm was adopted to transform the data into a more visually appreciable two-dimensional (2D) representation due to its fast computation, high accuracy and efficiency of processing non-linear data.

Fig. 5 demonstrates the comparison of cluster distributions of feature vectors as the number of clusters in the K-Means algorithm is increased. From the visualisation through the t-SNE technique, the aim is to identify one or more specific clusters with the indicative features' values, which could be used for generate the predictive rules.

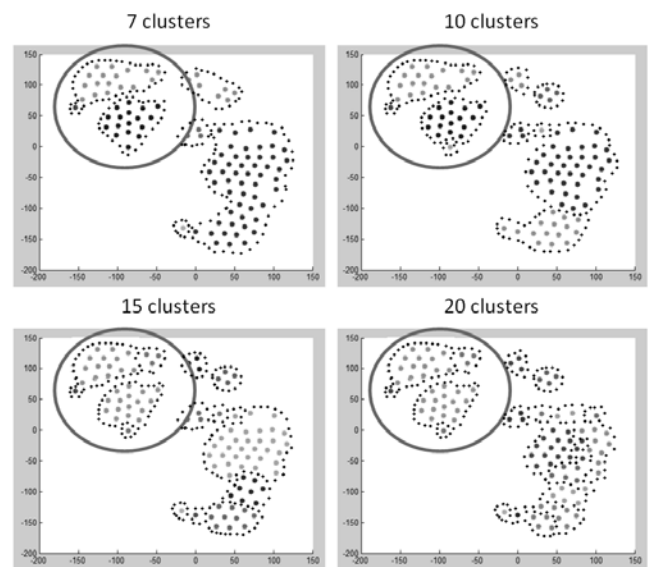


Fig. 5. t-SNE visualisation of K-Means clustering

As shown in Fig. 5, each cluster is separated with boundary lines. Each point represents one set of feature vector coordinates representing an instance of FP activity. The relative distance between points provides an indication of the similarity/dissimilarity between plotted feature vectors. The closer together these points are clustered, the greater the similarity that exists between associated FP activity. From the t-SNE visualisation, it is apparent that the data is distributed in two main areas, i.e. that inside and outside the circled area. While the membership of the clusters outside the circle changes significantly as the number of initialising clusters increases, the cluster membership inside the circled area does not.

It is evident that the distributed data in the circled area is more stable and consistent, and demonstrates a higher correlation coefficient. This means that the feature vectors within this cluster are most likely to lead to improved accuracy if used to build the predictive model. The purpose of the prognostic function is to predict potential future operation. Based on this analysis it is apparent that, if a new log file is presented to the KBS, rules operating on the NTP, ITT, ACA and APD features can be used to determine the TTT. That is, given the four feature values, the TTT (i.e. time to the next PMAR operation) can be predicted within a particular time window. This allows engineers to take actions before the occurrence of a fault causing PMAR operation.

C. Predictive Rule Implementation

In order to generate rules to predict future PMAR operation, Fig. 6 shows a parallel chart describing the distribution of values of the features characterising the cluster representing PMAR operation.

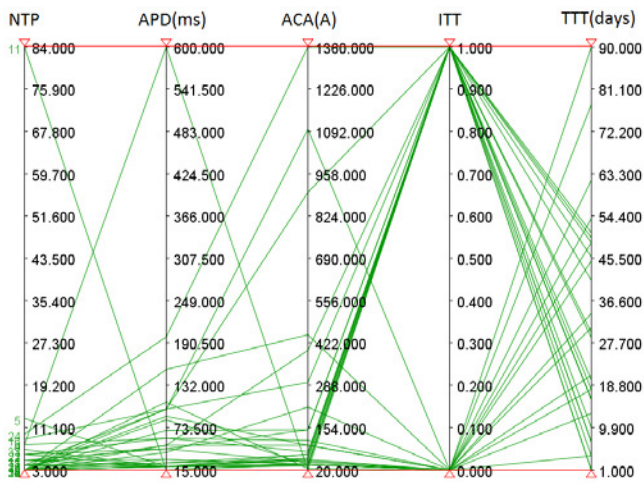


Fig. 6. Parallel coordination plot of the segmented data

The chart in Fig. 6 plots the range of values, and each continuous line represents the value of feature in a PMAR record. These are plotted for the data points which consistently remain members of the cluster shown as the circled area in Fig. 5. This chart indicates that the value range of the NTP, APD and ACA features are relatively concentrated. Therefore, this visualisation can be used to derive threshold settings for these features required to predict the TTT. Fig. 6 shows that this indicates a trip will occur within 3 months (90 days). However, there is noise within this data causing a problem in determining the accurate thresholds of the conditions for predicting TTT. Ordering Points To Identify the Clustering Structure (OPTICS) was used to remove the outlier (noisy) data points from the encircled data.

OPTICS is an algorithm used to segment data into clusters of varying density [27]. In this paper, it was used to detect and filter the outliers through the application of a reachability distance function within the OPTICS technique. Fig. 7 shows how the reachability distance function in the OPTICS methodology finds the outliers in the segmented data.

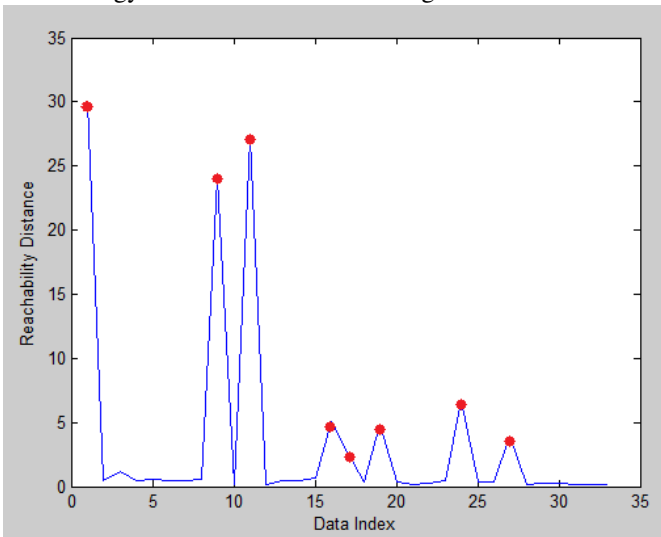


Fig. 7. Visualisation of noisy data with OPTICS algorithm

The function will automatically determine the cluster centre when clustering the raw data and obtain the distance between

the cluster centre and each data point. A higher reachability distance represents a greater distance from the cluster centre. Therefore, as illustrated in Fig. 7, 33 data points in the encircled clusters were analysed via the OPTICS technique to filter the outliers. The reachability distances of most data points are around 0, which means these data points are close to the cluster centre. However, there are eight data points with much higher reachability distances, which are marked as red points in Fig. 7. These eight data points are then removed as outliers. Consequently, after filtering this noisy data, the maximum and minimum value of defined features can be used to set the thresholds of the rule to predict the PMAR's operation. Fig. 8 demonstrates the new parallel coordination plot of the five features after the removal of outliers.

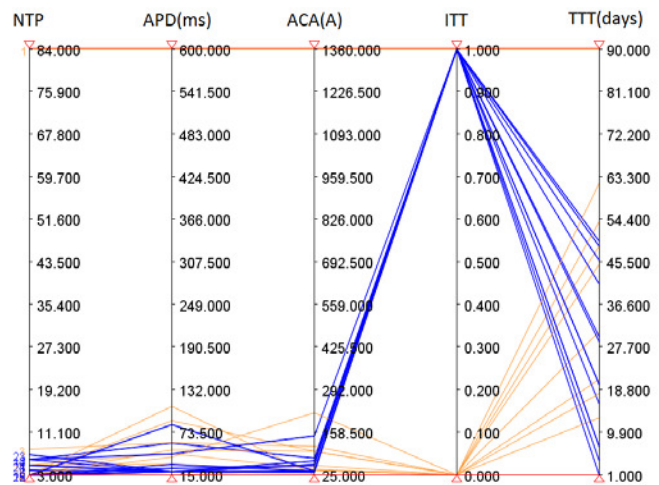


Fig. 8. Parallel coordination plot of features without noisy data

Fig. 8 shows the value ranges of the features NTP, APD, and ACA which form the threshold for the predictive rule shown in Fig. 9. Furthermore, Fig. 8 also shows that the ITT (i.e. the interval time between consecutive pick-ups) associated with the data in the circled cluster can be seen to both increase and decrease. The blue lines represent data where the value of feature ITT is True (i.e. the time interval between two consecutive fault pick-ups in a group of activity is decreasing), and the orange lines represent data where the value of feature ITT is False (i.e. the time interval between two consecutive fault pick-ups in a group of activity is increasing). From this observation the ITT feature is redundant in the classification of data in this cluster and consequently is not included in the derived rule in Fig. 9. The rule is also more precise in terms of the predicted time to trip. This process of generating a prognostic rule is repeatable. As more DA data becomes available, the thresholds in the rule would be more precise in terms of predicting future PMAR operation. The rule in Fig. 9 has been generated based on the actual network data available at present.

Predictive Rule:
 If A FP group is detected
 And 3 < number of pick-ups < 8
 And 15ms < average pick-up duration < 120ms
 And 25A < average pick-up amplitude < 220A
 Then
 There will be a trip within 2 months of the last pick-up activity

Fig. 9. The prognostic rule for predicting future PMAR's operation

D. Data Testing

To validate the rule, 27 unseen PMAR log files containing FP activities and PMAR operations were selected for analysis. Using historical data, it was found that there are 32 FP groups of fault pick-up activities that do not lead to any trips or lockouts within 1 month which ‘fire’ the rule in Fig. 9, suggesting that a trip will occur in 2 months. Through testing, it was established that 27 of these FP groups did indeed result in a PMAR operation within the predicted 2-month timeframe. Therefore, the accuracy or success rate of the rule 84.4%

This analysis of utility data verifies that the methodology adopted can be used to develop accurate predictive capabilities. Further predictive rules can be created using this approach as more data is collected by the network operator.

V. CASE STUDIES: IMPLEMENTATION OF THE VISUALISATION TOOL AND KNOWLEDGE BASES

Sections III and IV outline the data science methods and approaches to achieve diagnostic and prognostic functions in the DSS. Fig. 10 illustrates this, and shows that there is a data science based KBS development process, followed by its implementation within the DSS.

As shown in Fig. 10, the development of the KBS focuses on deriving and defining the diagnostic and prognostic rules, which are generated through the visualisation and data mining of actual PMAR historical data. It is also supported by expert knowledge and experience. In order to implement the functionality, a KBS is used to deploy the fault analysis rules. This has been developed using Drools Fusion [28] (a rule-based inference engine) to offer a flexible framework for processing the online PMAR data. It invokes the appropriate rules based on

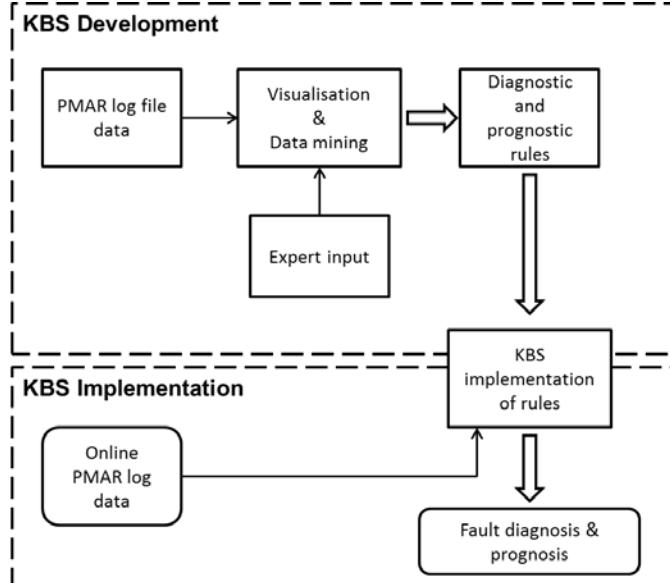


Fig. 10. Development and implementation of KBS

the input data for both the diagnostic and prognostic functions. The KBS is integrated within a prototype DSS. This was developed in a Java and included visualisations of results.

The case study is based around a PMAR which exhibited frequent supply interruptions. These frequent undiagnosed

events could be the result of the degradation of overhead lines and PMAR components. The PMAR log file was imported into the system for further automatic analysis. The following sections demonstrate the use of the functions within the system for fault diagnosis and prognosis.

A. Case Study Part 1: PMAR Device Fault Diagnosis

After importing the particular PMAR log file, the knowledge-based system will automatically identify PMAR device faults and generate a report through the DSS user interface. The diagnostic report contains the identified fault with a short explanation of the fault cause. For this particular PMAR, the report indicates: *Microswitch fault with the detection of an unexpected accelerated contact time.*

In order to validate the result of automated diagnosis of PMAR device faults, the original PMAR log file is selected and analysed. The unexpected accelerated contact time is based on the time interval between action signals and status changes being less than the initial setting. Fig. 2 shows the diagnostic rule and Fig. 11 the data that fires it. This Microswitch fault could lead to unnecessary PMAR operations, contributing to the number of short-term outages.

01/09/2010 04:01:23:585	Trip			OC1+	B
01/09/2010 04:01:23:585	Trip			OC1+	C
01/09/2010 04:01:23:601	Open			Driver	
01/09/2010 04:01:23:645	AR Initiation			AR OCEF	
01/09/2010 04:01:23:645	Pickup			End	OC1+ B
01/09/2010 04:01:23:645	Pickup			End	OC1+ C
01/09/2010 04:01:23:661	Pickup			Start	Uab<
01/09/2010 04:01:23:701	Pickup			Start	LSD
01/09/2010 04:01:23:701	Pickup			Start	Urst<
01/09/2010 04:01:33:643	Close			AR OCEF	
01/09/2010 04:01:33:672	Closed			Driver	
01/09/2010 04:01:33:703	Pickup			End	LSD

Fig. 11. Validation for Microswitch fault diagnosis

B. Case Study Part 2: Semi-Permanent Fault Detection

As well as the KBS automatically identifying the Microswitch fault, a user interface and visualisation tool was developed to provide the control engineers with detailed information on classified fault scenarios. Fig. 12 illustrates the classification of PMAR operation based on the affected phases.

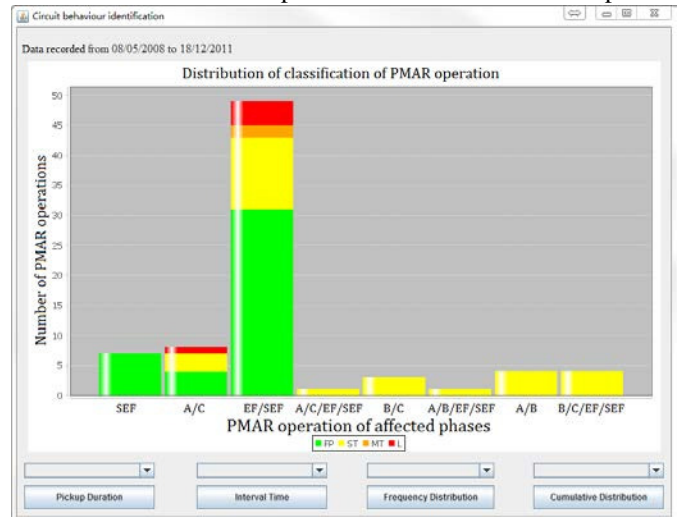


Fig. 12. Classification of PMAR operation

In Fig. 12, the different fault phases are listed with numbers and classifications of PMAR operation, marked by distinct

colours. Obviously, the Earth Fault and Sensitive Earth Fault (EF/SEF) is the most frequent of the eight categories of fault detected in this particular PMAR. The EF/SEF fault events contain 31 FP activities and 12 ST activities. These indicate that frequent short-term fault activities could be caused by a semi-permanent fault and lead to a permanent fault. At the bottom of the frame, the engineers can expand the details of the key features (i.e. DT, IT, FD, CFD) describing the selected fault types. Fig. 13 gives the detailed information of EF/SEF's FD and CFD.

With the frequency distribution shown in Fig. 13, the EF/SEF fault always disappears without any PMAR operation or is cleared by a single attempted reclosure. Meanwhile, the number of FP activities increase during the same time period. Though the FP fault scenario does not lead to any PMAR operations, the growing number indicates increasing severity of this fault condition. This may result in more supply interruptions or a long-term outage in the future. Furthermore, the cumulative distribution reflects an increasing rate of EF/SEF occurrences. According to these analyses, the engineers could conclude that the EF/SEF fault events could be caused by the existence of a semi-permanent fault.

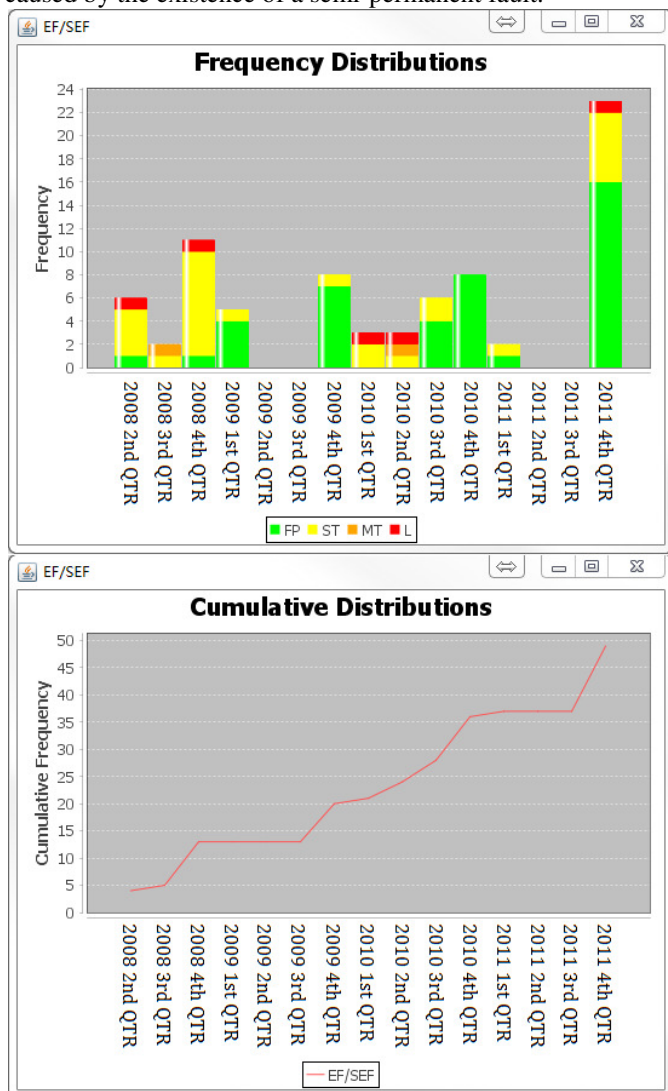


Fig. 13. The FD and CFD of EF/SEF

C. Case Study Part 3: Fault Prognosis

After diagnosing the PMAR device fault and detecting the semi-permanent fault, engineers can also utilise the KBS application to check the fault prognosis report based on the predictive rule developed. The KBS reports a potential PMAR operation will occur within 2 months from the date of 14/12/2011, and the system automatically informs the user of this.

Fig. 14 shows data from the PMAR log file. It can be seen that the time stamp of the next trip (which is highlighted) is 18/12/2011. This is within the predicted time period of 2 months from 14/12/2011.

18/12/2011 07:07:11:484	Pickup		Start	OC1+	A
18/12/2011 07:07:11:484	Pickup		Start	OC1+	C
18/12/2011 07:07:11:535	Toir		End	IR	
18/12/2011 07:07:11:629	Trip			OC1+	A
18/12/2011 07:07:11:629	Trip			OC1+	C
18/12/2011 07:07:11:669	Open			Driver	
18/12/2011 07:07:11:669	Pickup		End	OC1+	A
18/12/2011 07:07:11:669	Reset			OC1+	A

Fig. 14. Validation for prediction

VI. DISCUSSION AND CONCLUSION

This paper details the data analysis approach, techniques, methods and prototype development for a decision support system to automate fault diagnosis and prognosis of trips within distribution automation. It supports the reduction of customer supply interruptions and detection of underlying asset deterioration. The research contributions are:

- the design and proof of a predictive capability to identify emerging faults within a DA application;
- the combination of data science and knowledge based techniques to deliver diagnostic and prognostic functionality; and,
- a unique focus on distributed DA autoreclosing devices.

The research is also timely in that it addresses the use of low-voltage distribution network IED data for fault diagnosis and prognosis. The data analysis methods can be applied to wider distribution automation devices and the case studies demonstrate how the automatic system assists the control engineers with their fault data analysis. A flexible and effective knowledge-based system for identifying fault events and delivering prognostic functions was produced, and this can be extended over time and with greater data access. The utility is now in the process of initiating a project to roll this functionality out across all of their PMARs.

REFERENCES

- [1] V. Madani, R. Das, F. Aminifar, J. McDonald, S. S. Venkata, D. Novosel, A. Bose, and M. Shahidehpour, "Distribution Automation Strategies Challenges and Opportunities in a Changing Landscape," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 2157-2165, Jul. 2015.
- [2] R. Smith. (2015, Nov.). Electricity Ten Year Statement 2015. Nationalgrid. UK. [Online]. Available: <http://www2.nationalgrid.com/UK/Industry-information/Future-of-Energy/Electricity-Ten-Year-Statement/>
- [3] I. Lim, and T. S. Sidhu, "Design of a backup IED for IEC 61850-based substation," *IEEE Trans. Power Delivery*, vol. 28, no. 4, pp. 2048-2055, Oct. 2013.
- [4] S. McArthur, A. Dysko, J. McDonald, S. Bell, R. Mather, and S. Burt, "The application of model based reasoning within a decision support system for

- protection engineers," *IEEE Trans, Power Delivery*, vol. 11, no. 4, pp. 1748-1754, Oct. 1996.
- [5] S. D. J. McArthur, E. M. Davidson, V. M. Catterson, A. L. Dimeas, N. D. Hatzargyriou, F. Ponci, and T. Funabashi, "Multi-agent systems for power engineering applications—Part I: Concepts, approaches, and technical challenges," *IEEE Trans, Power Systems*, vol. 22, no. 4, pp. 1743-1752, Nov. 2007.
- [6] C. J. Edwards, E. M. Davidson, S. D. J. McArthur, I. Watt and T. Cumming, "Flexible Model-Based Alarm Processing for Protection Performance Assessment and Incident Identification," *IEEE Trans, Power Systems*, vol. 28, no. 3, pp. 2584-2591, Jul. 2013.
- [7] P. Balakrishna, K. Rajagopal, and K. S. Swarup, "Look-ahead distribution power restoration analysis based on integrated operation of distribution automation and advanced metering infrastructure systems," *IET Journal of Generation, Transmission & Distribution*, vol. 9, no. 14, pp. 2024-2031, Oct. 2015.
- [8] M. Kezunovic, "New Trends for Automated Fault and Disturbance Analysis," in *PAC World*, 2013, pp. 54-59.
- [9] T. Popovic, M. Kezunovic, and B. Krstajic, "Implementation requirements for automated fault data analytics in power systems," *Int. Trans. Electr. Energ. Syst.*, vol. 25, no. 4, pp. 731-752, Jan. 2014.
- [10] A. Angioni, J. Shang, F. Ponci and A. Monti, "Real-Time Monitoring of Distribution System Based on State Estimation," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 10, pp. 2234-2243, Oct. 2016.
- [11] T. Dekkers, M. Zut, and K. Gleeson, "A new approach to intermittent fault management in low voltage networks to reduce the number of repeat interruptions," in *IEEE Electricity Distribution Conf., CIGRE 2013*, pp. 1-3.
- [12] J. A. Wischkaemper, C. L. Benner, B. D. Russell and K. Manivannan, "Application of Waveform Analytics for Improved Situational Awareness of Electric Distribution Feeders," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 2041-2049, Jul. 2015.
- [13] T. Popovic, M. Kezunovic, and B. Krstajic, "Smart grid data analytics for digital protective relay event recordings," *Information Systems Frontiers*, vol. 17, no. 3, pp. 591-600, Jun. 2013.
- [14] NOJA POWER [Online]. Available: <http://www.nojapower.co.uk/product/recloser.html>.
- [15] J. Ren and M. Kezunovic, "Teaching Protective Relaying Design and Application Using New Modeling and Simulation Tools," *Journal of Energy and Power Engineering*, vol. 6, no. 5, pp. 762-770, May. 2012.
- [16] F. Mitchell. (2016). SP Transmission Annual Performance Report 2015/16. SP Energy Network. UK. [Online]. Available: http://www.spenergynetworks.co.uk/userfiles/file/SPEN_TransmissionReport.pdf.
- [17] C. Ly, K. Tom, C. S. Byington, R. Patrick and G. J. Vachtsevanos, "Fault diagnosis and failure prognosis for engineering systems: A global perspective," in *IEEE Automation Science and Engineering Conf., CASE*, 2009, pp. 108-115.
- [18] J. R. McDonald, G. M. Burt, J. S. Zielinski and S. D. J. McArthur, "Intelligent Knowledge Based Systems in Electrical Power Engineering," 1st ed., 1997, pp. 5-17.
- [19] I. K. Song, S. Y. Yun, S. C. Kwon and N. H. Kwak, "Design of Smart Distribution Management System for Obtaining Real-Time Security Analysis and Predictive Operation in Korea," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 375-382, Jul. 2013.
- [20] X. Wang, S. M. Strachan, S. D. J. McArthur and J. D. Kirkwood, "Automatic analysis of Pole Mounted Auto-Recloser data for fault diagnosis and prognosis," in *IEEE 18th Intelligent System Application to Power Systems Conf., ISAP*, 2015, pp. 1-6.
- [21] E. Frank and I. H. Witten, "Data Mining: Practical Machine Learning Tool and Techniques," 3rd ed., MK, 2011, pp. 47-70.
- [22] X. Peng, C. Zhou, D. M. Hepburn, M. D. Judd and W. H. Siew, "Application of K-Means method to pattern recognition in on-line cable partial discharge monitoring," *IEEE Trans, Dielectrics and Electrical Insulation*, vol. 20, no. 3, pp. 754-761, Jun. 2013.
- [23] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview," *The Journal of WIREs: Data mining and knowledge discovery*, vol. 2, no. 1, pp. 86-97, Dec. 2011.
- [24] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman and A. Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," *IEEE Trans, Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 881-892, Jul. 2002.
- [25] L. V. D. Maaten and G. Hinton, "Visualizing data using t-SNE," *The Journal of Machine Learning Research*, vol. 9, pp. 2579-2605, Nov. 2008.
- [26] S. M. Strachan, S. D. J. McArthur, B. Stephen, J. R. McDonald and A. Campbell, "Providing Decision Support for the Condition-Based Maintenance of Circuit Breakers Through Data Mining of Trip Coil Current Signatures," in *IEEE Transactions on Power Delivery*, vol. 22, no. 1, pp. 178-186, Jan. 2007.
- [27] M. Ankerst, M. M. Breunig, H. P. Kriegel, and J. Sander. "OPTICS: ordering points to identify the clustering structure," in *Management of Data Conf.*, 1999, vol. 28, no. 2. pp. 49-60.
- [28] M. Salatino, E. Aliverti and M. N. D. Maio, "Mastering Jboss Drools 6 for Developers," 1st ed., Packt, 2016, pp. 1-330.

Xiaoyu Wang (S'16) received the B.Eng. (Hons.) degree from the University of Strathclyde, Glasgow, U.K., in 2012, where he is a Ph.D. degree candidate in electronic and electrical engineering. His research interests include fault diagnosis and prognosis, distribution automation, and intelligent system application in power system condition monitoring and data analytics.

Stephen D. J. McArthur (M'93–SM'07–F'15) received the B.Eng. (Hons.) and Ph.D. degrees from the University of Strathclyde, Glasgow, U.K., in 1992 and 1996, respectively. He is now the co-Director of the Institute for Energy and Environment at the University of Strathclyde. His research interests include intelligent system applications in power engineering, covering condition monitoring, diagnostics and prognostics, active network management and wider smart grid applications.

Scott M. Strachan received the B.Eng. (Hons.) and Ph.D. degrees from the University of Strathclyde, Glasgow, U.K., in 1995 and 2005, respectively. Currently, he is a Teaching Fellow within the Institute of Energy and Environment, University of Strathclyde. His research interests include plant monitoring, asset management, data mining, knowledge management and engineering, and intelligent systems applications in power engineering.

John D. Kirkwood was Automation Manager with ScottishPower Energy Networks, Cambuslang, U.K., in their electrical distribution operations division. John sadly passed away during this research. He is missed by his colleagues, and this work is an example of his commitment to research, innovation and his ability to mentor and develop engineers.

Bruce Paisley is Senior Automation Engineer with ScottishPower Energy Networks, Cambuslang, U.K., in their electrical distribution operations division.