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

Xiaoyu Wang

A thesis submitted for the degree of Doctor of Philosophy to
the Department of Electronic and Electrical Engineering
University of Strathclyde

Aug 2017

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

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 thesis details the design of a novel decision support system to achieve automatic fault diagnosis and prognosis for DA schemes. It combines detailed data from a specific DA device with SCADA data, by utilising rule-based, data science techniques (e.g. 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 condition of device faults, 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 using prognostic functionality. The novel contributions also include the characterisation and identification of semi-permanent faults and a 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.

Acknowledgements

I would like to sincerely thank my supervisors, Prof. Stephen McArthur and Dr. Scott Strachan, who have been extremely supportive to my research. During the years, my supervisors for my growth pour a lot of effort and effort. Their guidance and inspiration are truly invaluable in all the time of research and writing of this thesis, as well as my life and future career. Their rigorous and pragmatic attitude and profound academic attainments have benefited me for life. Once again, I would like to extend my respect and thanks to my two supervisors.

Besides my supervisors, I'd also like to express my gratitude to my colleagues in the Advanced Electrical Systems group at Strathclyde, who have been really kind and helpful. Special thanks to Dr David Pattison. I'm really grateful for his insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives.

My sincere thanks also goes to ScottishPower Energy Networks for the financial and technical support to this research. I'd like to thank John Kirkwood and Bruce Paisley for their valuable contributions to my work. Without their precious support it would not be possible to conduct this research.

Here, I need to thank everyone who helped my research and life, and also thank all my friends for the time spent with me, for the sleepless nights we were working together, and for all the fun we have had in the last few years.

Finally, but by no means least, I'd especially like to thank my families and my fiancee for supporting me spiritually throughout writing this thesis and my life in general. This thesis is dedicated to them.

Contents

List of Figures

List of Tables

Glossary of Abbreviations

AC	Alternating Current
ACA	Average Current Amplitude
AI	Artificial Intelligence
ANN	Artificial Neural Network
APD	Average Pick-up Duration
AVR	Automatic Voltage Regulator
CB	Case-Based
CBR	Case-Based Reasoning
CCTV	Closed-Circuit Television
CFD	Cumulative Frequency Distribution
CI	Customer Interruption
CML	Customer Minute Lost
CT	Current Transformer
DA	Distribution Automation
DAS	Distribution Automation System

DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DER	Distributed Energy Resource
DFR	Digital Fault Recorder
DG	Distributed Generator
DL	Deep Learning
DL-G	Double Line-to-Ground
DNO	Distribution Network Operator
DPR	Digital Protective Relay
DSM	Demand-Side Management
DSS	Decision Support System
DT	Duration Time
EF/SEF	Earth Fault and Sensitive Earth Fault
EHV	Extra High Voltage
FD	Frequency Distribution
FP	Fault Pick-up
GDE	General Diagnostic Engine
GUI	Graphical User Interface
HMI	Human Machine Interface
HV	High Voltage
IE	Inference Engine

IED	Intelligent Electronic Device
IEEE	Institute of Electrical and Electronics Engineers
IT	Interval Time
ITT	Interval Time Trend
JESS	Java Expert System Shell
KB	Knowledge-Base/ Knowledge-Based
KBS	Knowledge-Based System
KNN	K-Nearest Neighbour
L	Lockout
LAN	Local Area Network
L-L	Line-to-Line
LLE	Locally Linear Embedding
LV	Low Voltage
MAS	Multi-Agent System
MB	Model-Based
MBR	Model-Based Reasoning
MPM	Main Processor Module
MT	Multiple Trip
MTU	Master Terminal Unit
MV	Medium Voltage

NOP	Normally Open Point
NTP	Number of Total Pick-ups
OC	Open Circuit
OHL	Overhead Line
OPTICS	Ordering Points To Identify the Clustering Structure
PBX	Private Branch Exchange
PCA	Principal Components Analysis
PD	Partial Discharge
PEDA	Protection Engineering Diagnostic Agents
PLC	Programmable Logic Controller
PMAR	Pole Mounted Auto-Recloser
PMU	Phasor Measurement Unit
Ph-Ph-Ph	Phase-to-Phase-to-Phase
PSN	Public Switched Network
RB	Rule-Base/Rule-Based
RTU	Remote Terminal Unit
SAIDI	System Average Interruption Duration Index
SAIFI	System Average Interruption Frequency Index
SC	Short Circuit
SCADA	Supervisory Control And Data Acquisition

SELF	Single Ended Location of Fault
SL-G	Single Line-to-Ground
SPEN	ScottishPower Energy Networks
SPF	Semi-Permanent Fault
SSC	Source Short Circuit
ST	Single Trip
t-SNE	t-Distributed Stochastic Neighbour Embedding
TTT	Time to Trip
TXT	text
UHF	Ultra High Frequency
UO	Unsolicited Opening
VT	Voltage Transformer
WAN	Wide Area Network

Chapter 1

Introduction

1.1 Introduction to the Research

Since the emergence of the smart grid paradigm, existing electricity grids around the world have seen a rapid increase in the deployment of new technologies. The increasing prevalence of microgrid developments, energy storage, and renewable energy has led to noticeable changes in the power system characteristics. In tandem, Distribution Network Operators (DNOs) are also focusing on improving the reliability and stability, operational resiliency, and customer service of the distribution systems [Sma14, MDA$^+$15, Far10, SM11c]. These activities are supported by Distribution Automation (DA) strategies, which guide the use of appropriate technologies to deliver the advances in reliability. DA therefore exerts a critical influence on the development of smart grids, especially distribution systems, when encountering potential challenges and opportunities in improving the planning, operation, maintenance, and protection [IA09]. For example, the planning and grid-connection of backup power resources to mitigate the sudden outage and restore electricity is based on the implementation and administration of DA strategy [MDA$^+$15].

In general, DA deployed in the distribution network is to reduce the interruption duration, isolate the faulted area, and rapidly reconnect customers following network faults. DA can also assist in power system planning by automating the closure of normally open points for optimal power flow [SM11c]. Planning, controlling, and maintaining distribution systems have become more complex as loads increase and regulatory regimes focus increasingly on the security of supply and reliability. This is therefore a driving factor to improve the protection and operation of those systems. In order to effectively manage the networks, major investments and developments have subsequently been undertaken in the area of DA [MDA+15, DMA+15].

The implementation of DA technologies has evolved through recent related technological developments. In the past, the restoration of electricity supply always required the control engineers or maintenance staff to manually operate the protection devices. However, new installations of fault circuit indicators or fault passage indicators (i.e. devices which provide an indication of circuit faults via visualisation or remote tools) offer more flexibility in the detection and isolation of faulted areas, instead of manual awareness and observation [DMA+15, VFS13]. Distribution systems are now widely equipped with multi-functional Intelligent Electronic Devices (IEDs) [KM10, Ibr12, PKK15]. Each device may have integrated one or more functions, such as protection, monitoring, communication, and data recording. When these functions are combined or integrated within the Supervisory Control And Data Acquisition (SCADA) system [BWM14, SM11c, Off04], the distribution system can be promptly restored when a transient disturbance occurs, or reconfigured following a long-term power outage. These automated operations, under modern DA schemes, help improve the system's reliability indices [Rel12, APP14, NVKH11].

The metering capability of the installed IEDs can be used to process status changes and parameters of the integrated protection and control elements when they experience a fault event or anomalous activity. This captured data can provide an explicit view of real-time network scenarios by transmission of the data through a communication system instead of manually logging and reporting the records from each item of DA equipment. The control engineers can utilise the data from SCADA and IEDs to analyse the underlying circuit's conditions to achieve a fault diagnosis, or even extract emerging fault information over a period of time to predict future behaviour. Both the fault diagnosis and prognosis could be achieved either off-line or on-line. Thus, the data analysis of the DA operations can support asset management, provide capabilities of reactive fault management and post fault investigation. It could also support potential predictive capabilities which identify the evolving faults and offers the opportunity to take pre-emptive action.

However, the ongoing implementation of DA reveals two issues. First, although the IEDs have a critical role in DA, it is not cost effective to install a large number of them, in view of the costs for purchase and maintenance. So the DNOs need to balance the costs of installation against operational value. Second, large-scale deployment of IEDs have led to increasing volumes of operational data that is becoming overwhelming for the manual processes currently used for data collection and analysis [SSLFF15]. For example, the IED data used in this research work is generated monthly within over 200 log files (each file contain about tens of thousands of sets of data, detailed in Appendix A). Therefore, in order to reasonably implement the DA technologies in the distribution systems and efficiently analyse the available network data, the only feasible way is to acquire and process the data automatically [McD03, SMB+10].

As a consequence, DA data is utilised for automatic fault diagnosis in various areas, including distribution generator and substation fault diagnosis, fault location identification, fault isolation and restoration, etc. [MR15, Che12, Che11, SdOF+08, THL14, GCF16, SRS+12]. Among the areas, many in the research community have focused on analysing the SCADA data and available network data (e.g. IEDs' data) for fault diagnosis to assist engineers to deal with distribution network events. Such systems provide diagnostic functions or decision support for power control and planning by interpreting SCADA data alongside appropriate Digital Fault Recorder (DFR) data, Phasor Measurement Unit (PMU) data or Digital Protective Relay (DPR) data [DMM+06, HMMM03, BMM+98, WVKW16, PKK15, DZG13]. However, the existing systems mainly utilise the available IEDs' data for supporting the analysis of SCADA data for protection validation and fault location or detection. For the root causes of fault events, there is no in-depth study. Furthermore, these systems are rarely designed and implemented to consider the use of data from intelligent DA devices for predictive functions.

The research reported in this thesis takes a further step and provides both diagnosis of DA equipment problems and also provides a prediction of potential faults and outages that are emerging. It is concerned with the design of a novel Decision Support System (DSS) that comprehensively and automatically analyses the captured data from the SCADA system and protection IEDs. The diagnostic and prognostic capabilities allow prevention of the outage taking thus positively impacting security and reliability. Based on the use of experts' experience, a Knowledge-Based system (KBS) approach has been developed and implemented. Such an approach is considered as the most suitable solution for addressing fault diagnosis and prognosis challenges.

In this thesis, the automatic DSS developed uses the KBS to diagnose the nature of a circuit's previous fault activity, identify underlying anomalous circuit activity, and highlights indications of problematic events gradually evolving into a full scale circuit fault. These functions are achieved by analysing the SCADA alarm data and Pole Mounted Auto-Reclosers' (PMARs') data (i.e. one specific type of protection IED), which has been provided by a network operator (ScottishPower Energy Networks (SPEN) was a partner in this research and co-funded it). The system has been developed to assist the analysis of SPEN's daily control report automatically generated every morning as part of SPEN's routine operations. The report summarises anomalous protection operations by filtering SCADA alarms in last 24 hours in order to prioritise maintenance scheduling.

Since the main purpose of the research work is to automate the manual processes of daily PMAR protection operation analysis using SCADA alarm data, and then identify the root causes of abnormal events on the circuits through analysis of PMAR data, so the DSS is fully designed to automate the entire process. The research drives towards the design of a fully automated process which provides the control engineers early indication of issues in a rapid and effective manner. The implemented KBS automatically analyses the data by invoking the Rule-Base (RB) generated from the expert knowledge, data mining and clustering techniques. The methodology is demonstrated through the design and implementation of this KBS. The key functions and their advantages are demonstrated through case studies based on actual network configurations and operational data captured from SPEN's distribution network.

In addition, the DSS provides automated data analysis and prediction, and builds this from visualisation tools developed as part of the research. These visualisation tools also offer the end user advanced functionality when the system is implemented. The benefit of such a visualisation tool is evaluated and demonstrated through the case studies.

1.2 Justification for Research

Research concerning the utilisation and analysis of SCADA data and DFR data has been reported since the early 90s [KRFS94, MMBB95]. This research focused on protection performance analysis and validation, based on expert system technology. The Rule-Based (RB) system was designed to provide diagnostic information by taking DFR data into consideration in 1992 [SLC$^+$92]. With the increasing metering capabilities integrated with the protection IEDs, the data produced by them becomes more worthy for automatic fault analysis, including calculation of fault distance and resistance [IRS07], protection operation validation and diagnosis [LK05], identification of fault location and operational decision support [DZG13, PKK15, GR13, GP13]. However, the data analysis research referred focused on diagnosing the recorded fault events for operational validation and decision support to assist engineers. In contrast, the PMAR is one category of protective IED which records the data that can indicate underlying and emerging anomalous activities as well. These activities are different from fault events, do not result in protection operations, but are closely related to and often precede them. Hence, this research focuses on the development of a DSS with analysis of both fault events and the data that indicates anomalous activities prior to the fault. It delivers fault diagnosis and prognosis through the implemented KBS.

RB systems are KBS where the experts' knowledge has been codified into logical rules for automated use and analysis. These rule-based KBS have been extensively used in power system applications, including power system condition monitoring and event classification [SBG02, SRM+08, RMJ10b], protection failure and event diagnosis [PWN92, MDM+96, BMM+98, OAA+10], service restoration and planning [KV91, PL97, SKK+96, AMG+97], and load forecasting [HHC+90, KEDH02b]. The successful applications of such knowledge-based expert systems have automated arduous tasks and provided technical support for engineers. Therefore, they offer an obvious solution to evaluate for the DA challenge in this research. The benefits of KBS are: engineering knowledge can readily be converted into appropriate rules; and, KBS can also provide excellent explanation facilities. This is extremely useful for identifying faults and suggesting the appropriate decision support for system operators.

Through comprehensive analysis of the decision support required around DA, this thesis defines the challenges, opportunities and benefits associated with automatic fault diagnosis and prognosis for distribution automation. This spans from the data manipulation to the implementation of the automated system. Adoption of the methodologies and techniques developed can lead to a more efficient and convenient process of analysing the conditions of circuits and intelligent DA devices for control engineers. They can be generalized for other DA devices and the more general analysis of distribution automation and network monitoring. For example, the methods and approaches presented in this thesis for analysing the fault activities captured from the PMARs could be applied for other similar protection IEDs installed in the distribution systems.

1.3 Principal Contributions

This research provides the following contributions to knowledge:

- Comprehensive investigation of AI techniques that are suitable for fault diagnosis and prognosis within the analysis of SCADA alarm data and PMAR data for improved network operation and reliability.

- The design and implementation of a DSS that employs a KBS for automatic analysis of fault events and anomalous activities, and demonstration of its operation through network case studies.

- The design and implementation of a visualisation tool that allows engineers to manipulate detailed DA log data, explore potential issues and obtain the results of the diagnostic and prognostic functions.

- The use of a data mining and clustering methodology to uncover predictive rules for indicating future potential network faults, and the implementation of these within a knowledge-based system.

- A unique focus on DA auto-reclosing devices and the identification of evolving or incipient network faults; and, the full implementation of a prototype with the design of specifications for an end-to-end automated data analysis system to support control engineers.

1.4 Thesis Overview

This thesis is organised as follows:

Chapter 2 provides an overview of Distribution Automation (DA), including its role in distribution systems, the main intelligent DA devices' functions and their fundamental principles, and the specific key issues and considerations associated with research work in fault diagnosis and prognosis. Existing activities and tools that are associated with DA fault diagnosis and prognosis are reviewed and discussed.

Chapter 3 reviews a number of AI techniques and their applications in automatic fault analysis for DA, based on which a knowledge-based approach is proposed for the diagnostic and prognostic tasks. The reasons of why such an approach is selected are discussed.

Chapter 4 presents the current manual process and outcome of the fault investigation with available network data, and the challenges around the move from a manual process to the automated DSS, with the details of the proposed knowledge-based methodology through description of the design and operation of the DSS.

Chapter 5 presents case studies using actual network data to demonstrate how the KBS can be used for fault diagnosis of PMAR device faults and circuit conditions, and how the designed visualisation tool integrated DSS system can offer the decision support for assisting engineers' diagnosis.

Chapter 6 presents the data mining and clustering methodology to define the predictive rules. Case studies of actual network data are used to demonstrate how the KBS can be used for fault prognosis, indicating future PMAR operations.

Chapter 7 summarises the work presented in the thesis, highlights the key challenges addressed and the contributions of the research. Future work is outlined which would extend and augment the system, and allow more widespread implementation.

1.5 Publications

The following publications have been completed during the course of this Ph.D.:

1.5.1 Journal Article

A Data Analytic Approach to Automatic Fault Diagnosis and Prognosis for Distribution Automation
X. Wang, S. D. J. McArthur, S. M. Strachan, J. D. Kirkwood, and B. Paisley
accepted by *IEEE Transactions on Smart Grid*, PP(99): 1-9, 2017. Available as IEEE Xplore early access: doi: 10.1109/TSG.2017.2707107.

1.5.2 Conference Papers

Decision Support for Distribution Automation Data Analytics for Automated Fault Diagnosis and Prognosis

X. Wang, S. D. J. McArthur, S. M. Strachan, and B. Paisley
International Conference on Electricity Distribution (CIRED), Glasgow, United Kingdom, 2017

Automatic Analysis of Pole Mounted Auto-Recloser Data for Fault Diagnosis and Prognosis
X. Wang, S. M. Strachan, S. D. J. McArthur, and J. D. Kirkwood
Intelligent System Application to Power Systems (ISAP), Porto, Portugal, pages 1-6, 2015. doi: 10.1109/ISAP.2015.7325519.

Automatic Analysis of Pole Mounted Auto-Recloser Data for Fault Prognosis to Mitigate Customer Supply Interruptions
X. Wang, S. M. Strachan, J. D. Kirkwood, and S. D. J. McArthur
International Universities Power Engineering Conference (UPEC), Cluj-Napoca, Romania, pages 1-6, 2014. doi: 10.1109/UPEC.2014.6934653.

Chapter 2

Review of Distribution System Protection and Automation

2.1 Introduction

This chapter reviews the fundamentals of distribution system protection and automation. It focuses on distribution automation strategy, implementation and the performance improvement through auto-recloser protection schemes. It starts with an introduction to electrical power systems detailing the characteristics of distribution systems and distribution automation in Section 2.2. Section 2.3 provides an overview of the SCADA systems supporting distribution automation, which covers the description of architecture and functions. Information on SCADA alarm data has also been provided to assist further analysis in the research work. In Section 2.4, the different electrical faults and main protection schemes and their functions are introduced, along with discussions of pole-mounted auto-reclosers in protecting the supply service in the distribution networks. The performance of distribution systems is introduced in Section 2.5, which classifies the electrical faults into supplementary categories and describes the reliability indices for evaluating performance. In Section 2.6, a review of existing activities associated with distribution automation fault diagnosis and prognosis is presented, and the value of analysing the SCADA alarm data and

IED data to provide fault diagnosis and prognosis in distribution automation are discussed.

2.2 Distribution Automation

To achieve a good understanding of Distribution Automation (DA) and its role in an electrical distribution system, a basic description of the main concepts of electric power systems is given in this section followed by the more details of distribution systems. The characteristics of a Distribution Automation System (DAS) are discussed, including the benefits and functions of DA technology. To enhance the brevity and relevance of this section, only selected schemes or applications of distribution network control and protection that are related to this research work are covered. For a comprehensive coverage of operational architectures and implementation strategies of DA further information can be found in [NGW07].

2.2.1 Electrical Power System

The aim of an electrical power system is to deliver electrical energy to consumers [SM11a]. The key systems are: generation system, transmission system and distribution system, which are shown in Figure 2.1.

The *generation* system is owned by one or several electric utilities who produce the electricity within power plants, and there is now an increasingly diversified range of renewable and low-carbon energy resources in use (e.g. solar energy, nuclear energy, wind power, hydro energy, and tidal energy, etc.) instead of traditional fossil fuel resources. Recently, Distributed Generators (DGs) are embedded into the power networks to satisfy local demand, but can also provide ancillary service in response to supply disturbances, that improve system reliability and performance [Mom07].

The responsibility of the *transmission* system is to deliver the power to the transmission and zone substations (i.e. primary and secondary substations in the UK) through the transmission lines. Due to the generated power being trans-

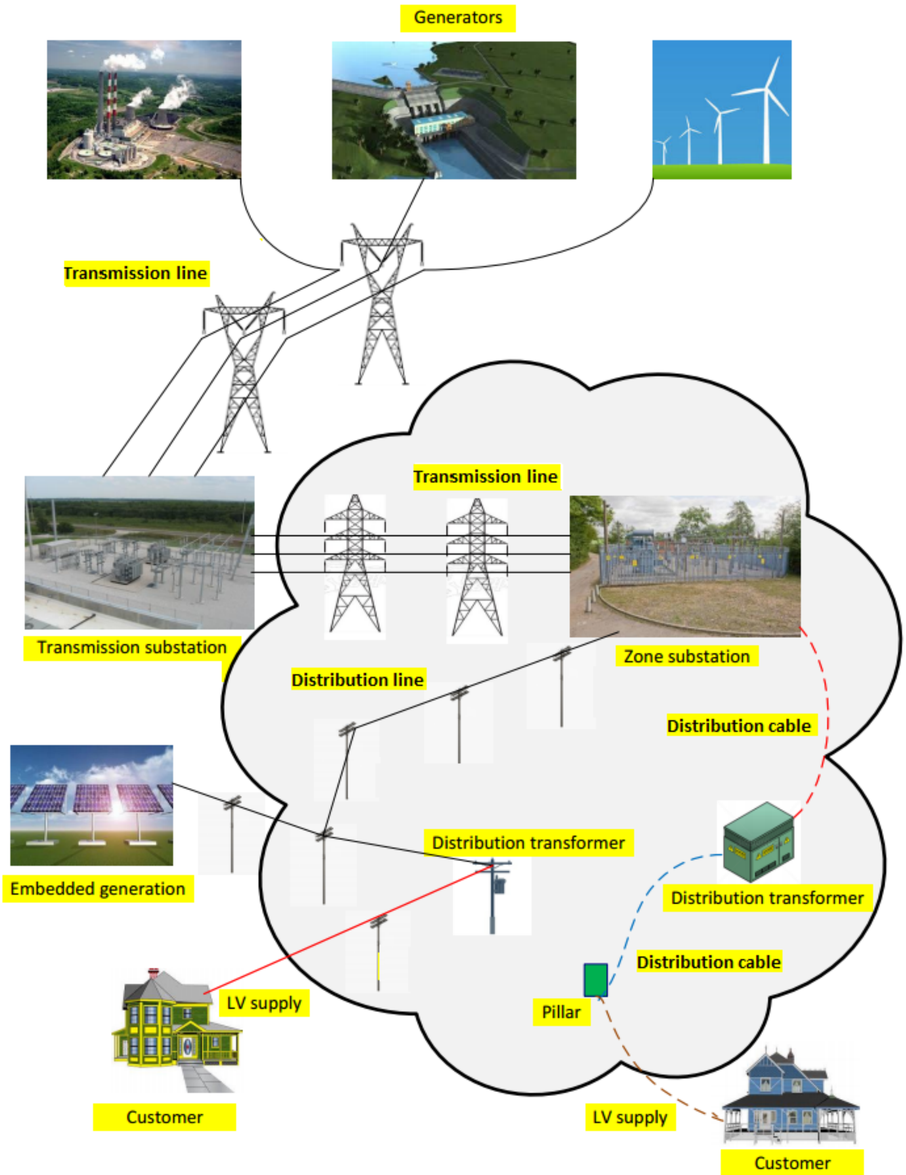


Figure 2.1: Electricity supply system [Act15]

mitted over long distances, the generation voltage level is increased by using the step-up power transformers as it reduces losses, where the major part of the energy losses comes from Joule effect in transformers and power lines (In general, average values of power losses at transmission and distribution levels are respectively 4%-6% and 4%-8%) during power transfers [IEC07]. The power voltage will be operated at Extra High Voltage (EHV), High Voltage (HV) and Medium Voltage (MV) level (typically 400 kV, 275 kV and 132 kV respectively in the UK) to the distribution networks.

In the *distribution* system, the power is first transmitted from the substations to the distribution transformers to obtain a stepped down voltage level, and the electricity is then ultimately distributed to the local consumers with appropriately Low Voltage (LV) levels through the distribution lines (or cables). For example, in the UK, the voltage level for industries and commercial establishments ranges from 11 kV to 33 kV, and the supplied voltage level of residential customers is at 230 V [Hat15].

2.2.2 Distribution System

As mentioned in 2.2.1, the distribution system is the final power link which connects the electric energy to the end users. The outlined area in Figure 2.1 is an instance of distribution systems. Alternating Current (AC) distribution is broadly deployed in present electric power networks [MM05, Ker01].

Commonly, the AC distribution system is divided into primary distribution network and secondary distribution network [BH12]. The primary distribution receives the bulk electric energy from the distribution substation and sends it to the distribution transformers to reduce the voltage level, or directly provide it to the large consumers. Afterwards, the secondary network distributes the suitable voltage level of electricity to lower wattage users (e.g. homes). Figure 2.2 shows the main components built in a distribution system: feeders, transformers, circuit breakers, voltage regulators and sectionalisers [Wil04].

The feeder is a conductor which transfers the electricity without tappings, the main factor of designing a feeder is the capacity for carrying the current. An

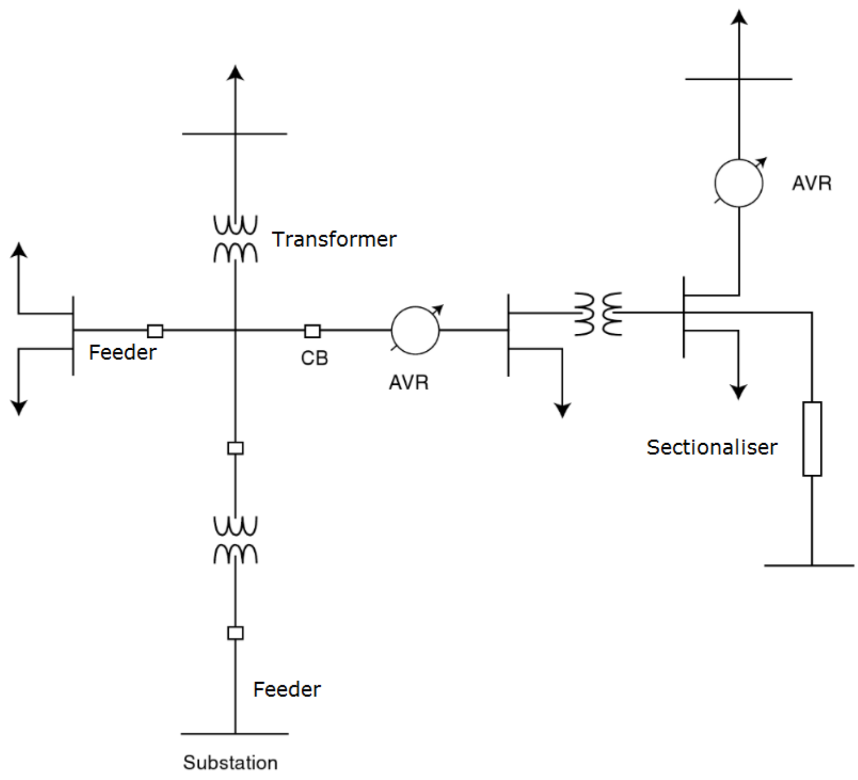


Figure 2.2: Distribution system [Mom07]

Automatic Voltage Regulator (AVR) in the distribution network maintains the stability of output voltage, particularly with respect to the connection of DG. The circuit breakers and sectionalisers are devices protecting the operation of distribution systems by automatically disconnecting the faulted equipment or areas. Certainly, a completely modern distribution system also contains other protection and control pieces, such as relays, auto-reclosers, sensors, etc. [Mom07]. Overall, these devices are installed with the purpose of improving the performance of distribution networks. The following section 2.4 will provide more details of protection devices on the distribution systems.

In order to ensure the high-quality and cost-effective power flow to the consumers, the topology and structure is a significant aspect in the design of a distribution system. Typically, the topology of a distribution system can be categorised into three types: radial, loop and mesh, as shown in Figure 2.3. A radial distribution network transmits the energy from a centre point to the branched customers, the power flows like water absorbed from the root of a tree to its branches. For a

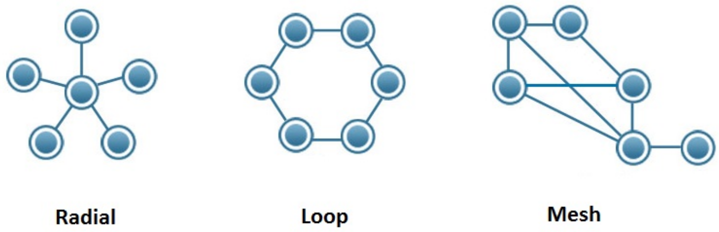


Figure 2.3: Different distribution network topologies

loop structure, there is a Normally Open Point (NOP) switch between the radial branches at the end of each feeder. When a fault is isolated at the upstream of a feeder, the downstream consumers can be restored by closing the NOP that leads to the power being delivered from another route. The mesh distribution system means the network is more highly interconnected with more available open points to reconnect the consumers when a fault occurs [Bus13]. Although the loop and mesh structure could increase the reliability of a distribution system, the complexity of operation will increase as well. Hence, appropriate communication and automation techniques are required to support the system reconfiguration and restoration instead of traditional control and protection mechanisms in the power networks.

2.2.3 Distribution Automation System

Nowadays, electricity distribution utilities have been striving to provide satisfactory reliability and power quality whilst efficiently managing their businesses. Therefore, in order to improve system reliability, improve operations, and offer better asset management, DA schemes are being designed and implemented by more and more distribution companies in many countries. For example, Figure 2.4 demonstrates the rapid increase of the number of automated switches (remote controlled) implemented outside the primary substations in the UK's distribution networks over a 9-year period (With the gradual maturity of the DA technology, more advanced single automated switch can be used in place of several 'older' versions hence leading to a reduction in the number of automatic switches deployed from 1999 to 2003). Experience shows the DA implementation improves reliabil-

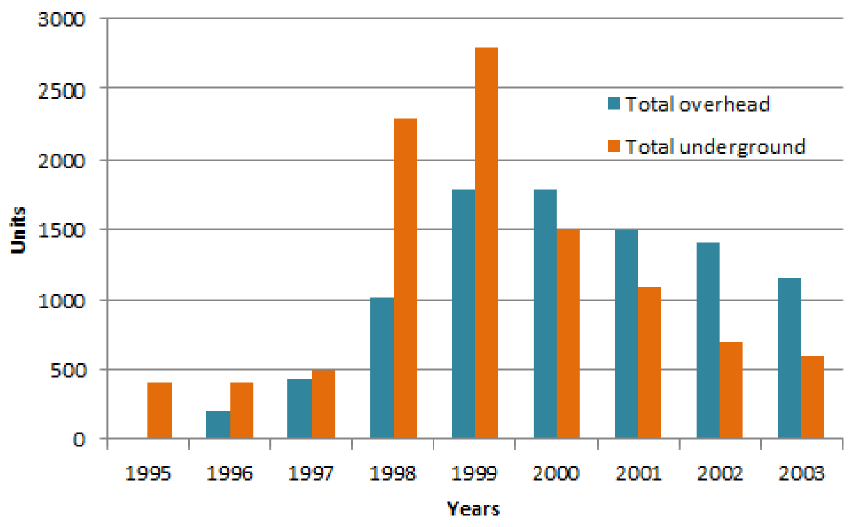


Figure 2.4: Number of switching devices installed per year outside the primary substations automated in the UK's distribution networks [NGW07]

ity by 20%-30% by reducing the number of outages and interruption durations in the distribution systems [NGW07].

There are many concepts and perceptions of the DA term. To put it simply, DA applies the automation into entire distribution system operation with associated information and communication technology applications. It is a cohesive architecture in power distribution systems, mixing together automation of local devices and central decision making. The structure normally consists of communicating relays, remote controlled switches, SCADA system, and distribution management and information processing system.

The DA functions improve the system performance during normal and abnormal situations automatically and efficiently [GASD11]. DA functions cover a wide range from fast isolation of the fault to the comprehensive consideration of control and planning, Figure 2.5 indicates the overall structure of effective DA functions. The value of these functions can be classified as follows [Mom07]:

- *Efficiency*: DA can minimise the power losses through the network restoration and reconfiguration by appropriate circuit switching for optimum load performance during an overload. And DA can reduce the energy usage and demand during peak times through the Demand-Side Management (DSM) analysis.

- *Reliability*: Reliability is one aspect of power quality issues, but it is considered separately with a greater focus on network outage situations. DA can reduce the duration and amount of supply outages through the quick system restoration and maintenance delivery. Meanwhile, DA could mitigate the potential power outages with reliability assessment or asset assessment by analysing historical recorded failures.

- *Quality*: In addition to reliability, power quality contains voltage sags and regulation, harmonic contents, etc. [KD11]. Intelligent DA devices could monitor the power quality and enable the dynamic controls, such as voltage regulation being automatically controlled through adjustments of capacitor banks and voltage regulators.

- *Security*: Physical plant security, cyber security and privacy protection become an important part of the modern distribution network, which could be implemented by applying DA technology. For example, the physical security of substations can use a CCTV systems to monitoring abnormal human activities integrated with a DA scheme, which may help keep a safe environment for stable system operations through allowing vision sensors to detect and analyse intruders [XLS+15].

The key benefits of DA for each of the categories (utility, network, and customer) are related to the divisions of efficiency, reliability, quality, and security. Firstly, DA can reduce operation and maintenance costs. For instance, fast fault location and isolation replace the staff dispatched and manual local operations, real-time data analysis of an asset management system provides an optimal maintenance plan before outages. Secondly, DA can improve the reliability and quality through the automatic control of installed intelligent devices responding to unexpected events. Thirdly, the improved information system of a DA offers more visibility for the engineers to plan and manage the networks to achieve their business objectives [NGW07, SW10, IEE07].

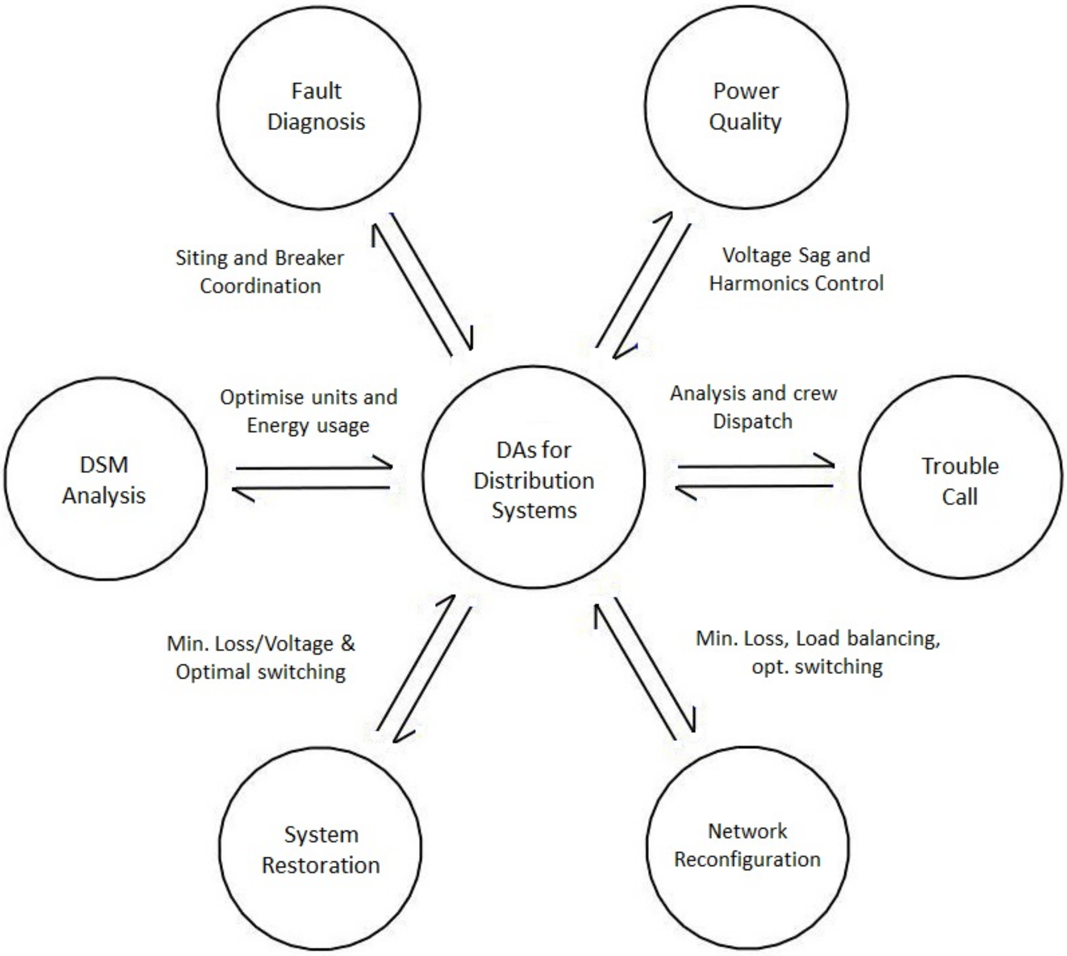


Figure 2.5: DA functions and structure [Mom07]

2.3 The SCADA System

Supervisory Control and Data Acquisition (SCADA) is the basis of any real-time power system control. A SCADA system is normally used to control and monitor the system in industries, such as electric energy, transportation, and renewable resource management. It acquires and pre-processes the data from different categories of intelligent devices (e.g. Remote Terminal Units (RTUs), Programmable Logic Controllers (PLCs)), and transfers it to the control centres or operator terminals to analyse and support decision-makings for system operations [Off04, Kum10, GD87]. The SCADA system can be complex (e.g. monitor all events in a power system) or relatively simple (e.g. control the temperature service in a building) depending on where it has been applied. Early traditional SCADA systems used Public Switched Network (PSN) for monitoring and control. In recent years, the communication of SCADA systems use corporate Local Area Networks (LAN) and Wide Area Networks (WAN), and some applications of wireless network technologies are integrated with the SCADA system seamlessly [Ale02, IEE93, McD93, KDS10]. In this thesis, the focus is on introducing the architecture and functions of a SCADA system, and detailing the SCADA alarm data which is analysed in the research work.

2.3.1 Architecture and Functions

The infrastructure construction of SCADA systems will encompass large and diverse equipment with the capabilities of monitoring, control, and communication. Generally, the structure of a SCADA system consists of four main components as follows [SM11c, Off04, AONI12, KC11]:

- Remote stations (i.e. RTUs or PLCs) deal with data transfer between field instrumentations and SCADA master stations. The field instrumentation is the device which measures the parameters on the plant and executes control operations with the notification from master stations. Here, remote stations gather converted data (e.g. analogue to digital data) recorded from field

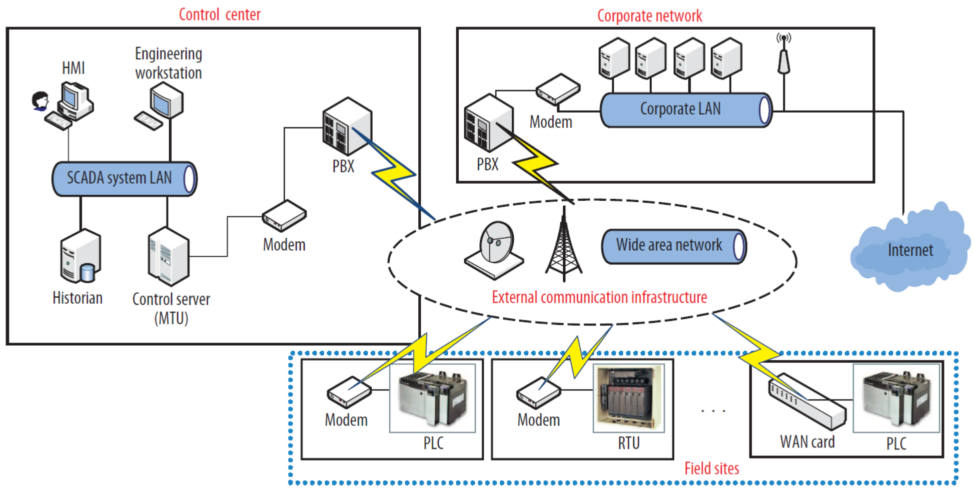


Figure 2.6: Simplified logical view of a typical SCADA architecture [AONI12]

instrumentations, and transmit it to the master station. On the other side, they receive the control commands from master stations and signal the field devices for operations.

- Communication networks are used to transfer data among the equipment in the whole SCADA system, including field devices, control units, and the SCADA central host. The channels of communication could be radio, leased cables, optical fibre, satellite, and so on. The configuration of communication networks depends on the size of the SCADA system, the number of local control units, and the rate of data update.

- Central control station is the heart of a SCADA system, sometimes it is called Master Terminal Unit (MTU) as well. It collects and stores data for processing and making necessary decision support, and it exchanges information with other systems through communication. Finally, the MTU provides results or proposed actions to the operators.

- Human Machine Interface (HMI) software systems are the bridges to connect the MTUs with the operators, they support the communication and allow the feedback (remote control) to the local field devices.

Figure 2.6 shows a basic view of a typical modern SCADA system's architecture which contains the mentioned main components, including the extended communication devices (modems and Private Branch Exchanges (PBXs)). In reality, there are three generations of SCADA architectures [SP 04, Sch12]:

- The 1^{st} Generation system (1970s technology) is "monolithic" [Off04]. At that stage, the SCADA system had no connectivity to other centralised systems. Meanwhile, the communication protocols were not feasible to transfer diverse categories of data between RTUs on the network, because of the protocols were generally provided by the system vendor. This led to the limited communication among master stations, remote stations, and the field devices.

- The 2^{nd} Generation (1980s/90s technology) The improvement of this generation was applying the LAN technology into distributing the data processes across multiple operation stations. Each station owned a specific function with sharing real-time information by connecting to the LAN. These operation stations could serve as communication processors and operator interfaces for communicating field devices with RTUs and providing visual assistances respectively. But the LAN connectivity between remote stations and the SCADA master station were still limited due to the RTU protocols, which were not available for other types of network traffic.

- The 3^{rd} Generation (2000s technology) is the networked SCADA system, the current generation architecture was widely deployed from 1990s. It is similar to the second generation, the major advantage of the third generation is that the open standards and protocols allow the distributed processors across both the WAN and LAN. The WAN protocols bring the communications between remote stations and SCADA master station.

At present, due to the complexity and vast capital investment for upgrading the generation of SCADA systems in large utilities such as power, oil, and gas, quite a few industries are in a transition period altering from the 2^{nd} generation. For example, in the UK, current electrical distribution networks are mixed of 2^{nd} and 3^{rd} generation SCADA systems [SP 04]. However, the standard functions of the SCADA systems will not change regardless of the generations. For a modern SCADA system settled in the DA of distribution systems, it should have the following basic features [SFK06, Kir14, TM15]:

1. Data acquisition: the responsibility of a SCADA system in a distribution network is to collect and process the basic operation state information for supporting system control and protection. The information is automatically collected and transmitted to SCADA databases through the various installed RTUs or intelligent DA devices with communication ability. The data can be categorised into three types: status values, measured values, and energy values. The status values represent digital alarm signals reflecting the status of switching devices (i.e. contacts' closings and openings). The measured values mean the time-varying quantities, such as current, voltage, and power factor. These data will be measured at a fixed frequency (e.g. every 10 seconds, every 1 hour) based on the requirement. Both the digital and analogue measured values will be transformed and normalised before sending to the databases.

2. Event generation and processing: the collected real-time data has little information by itself. In order to evaluate the conditions of power systems and provide feasible actions or decision supports accurately, the SCADA systems need to monitor the presented data and compare them with normal values or limits stored in the historical databases. For the status monitoring, the received triggered alarm data could be a normal condition when reviewing the settings and previous assigned data. Without the monitoring, this alarm signal could generate an unnecessary event processing. For the limit or trend monitoring, the SCADA system always has a delay function. It stores the present measured data in temporary memory and combines it with the following detected changes for monitoring. If the trends or values exceed the default thresholds, the data will generate an event processing, or be treated as normal conditions. Event processing is a significant function in the SCADA systems because the processed alarms assist the operators'

decision makings that influence the real-time performance the first time. Usually, both the unacknowledged and identified alarms (with causes of event occurring) are listed for operators.

3. Control: the plant and equipment could be controlled manually by the operators and automatically from software applications through SCADA systems. For corresponding to the two types of monitoring, the manual operations from the control centre have two classes. One is direct command controls to the individual equipment, such as open or close of one switching device. The other one is control messages regulate the predetermined limits of the equipment, such as raise or lower the tap changer position. Meanwhile, the automatic control of the SCADA systems is based on the condition of the processed event invokes the pre-set rules.

4. Data storage: as stated previously, the collected data will be stored in the real-time SCADA databases for applications to process. The data captured from RTUs will overwrite the old values with the new ones when the database server receives the data. For the measured values, only the changed parameters will be updated in the databases. Sometimes, the real-time SCADA databases have the connection with external data warehouses, where the archived data could be utilised for further data mining, calculation and analysis.

In practice, the SCADA systems also contain the functions of decision support and reporting for control operation through the HMIs [Ver16]. The HMIs will present a comprehensive view of the network's conditions with reported specific information and summarised historical data trends. Additionally, the processed results and listed alarms helps the operators perform the tasks for managing the systems. On the whole, SCADA technology and its functions are indispensable for the large utilities which intend to control and communicate automatically with their distributed equipment in real-time.

2.3.2 SCADA Alarm Data

SCADA systems contain alarm signals with important network information. The formats of SCADA alarms vary greatly within the applications between different utilities. However, they generally indicate to the operators network disturbances monitored by intelligent devices. The alarm information can be displayed in a summary report or in lists of related events. Typically, the SCADA alarms are triggered to warn the operators to take actions against emergency situations. And sometimes, alarms are generated to inform that the disturbances are addressed automatically by SCADA systems themselves [CBG14, AS09].

Due to the SCADA alarm data can provide abundant information of real-time system conditions, the SCADA alarm data processing therefore becomes an outstanding aspect of diagnosing the abnormal events to improve the system performance across various domains [ZL98, LPS$^+$13], such as SCADA alarm processing that assists real-time water management by supporting decisions to the control centre [Agu14]. In particular, in the areas of power systems' generation, transmission, and distribution, SCADA alarm processing has been used to diagnose fault events or anomalies [ZBWD05, CQF$^+$11, TG13, KSE$^+$00]. The related activities and application of SCADA alarm data in power system will be detailed in section 2.6.

In this research work, the analysis of SCADA alarm data supports a portion of the whole designed automatic DSS for fault diagnosis and prognosis. And the SCADA alarm data is provided by one of UK's distribution network operators (SPEN). Similar to the usual alarm data produced by other SCADA systems deployed in electricity distribution networks, the provided alarms mainly describe the abnormal situations and feedback on the control actions (manual or automatic). From the view of lists of alarm data, some alarms are triggered by independent incidents, while others may be raised on the repeat occurrences of a particular fault in the network. Therefore, if the related alarms are detected and grouped together, these remaining undiagnosed events could be analysed to prevent potential disturbances resulting from the same fault condition in the future.

The analysis of SCADA alarms in this thesis focuses on identifying the frequent supply interruptions for a specific circuit or PMAR by detecting the real-time data stored in the PSALERTS database [SP 15]. PSALERTS (named by SPEN) is a data set that archives SCADA alarms which include the status information (e.g. the opening or closing of a PMAR, test control of a circuit breaker, etc.) associated with the particular circuit details:

- Activity log time, a synchronised time stamp, through the GPS clock by communicating the RTU integrated with local intelligent devices upon execution of the necessary operations automatically or controlled remotely.
- Circuit name, contains the names of particular circuits, PMARs, circuit breakers, and switching gears which trigger alarms.
- Status information, represents the contacts' status (i.e. closed or open) of related protection devices.
- Tripping information, describes the protection operation as either telecontrolled manually or tripped automatically by devices themselves.

With the analysis of these prominent features in the PSALERTS database, there is a potential means of identifying repeated PMAR's operation and status changes of problematic circuits. Then, the in-depth diagnosis and prognosis through analysis of the PMAR's data could allow intervention to prevent the outage taking place. It should be noted that there is no global time reference for the local intelligent devices (e.g. PMARs), the real time stamps of PMARs' operation could be different. This needs to be overcome in the system by retrieving the synchronised time stamps of recorded alarms in PSALERTS.

2.4 The Protection System

It is inevitable that electrical faults occur during the operation and control of the electricity systems. The impacts can cause the deterioration and breakdown of electric equipment, which could directly lead to the loss of life and equipment.

Consequently, any detected faults must be quickly isolated or removed from the power system. Therefore, the duty of a protection system is to minimise damage to the overall network by defending against the fault conditions to ensure the cost of coordination as low as possible [ES13, Blu07]. This section will firstly introduce the types of electrical faults usually experienced on the circuits, then describe different protection schemes with associated protection devices, and finally detail PMAR's protection mechanism with its related data analysed in this research work.

2.4.1 Electrical Fault

Faults in the power network always lead to abnormal voltage and current. Depending on the existence of current flow when a fault occurs, the electrical faults can be categorised into Open Circuit (OC) faults and Short Circuit (SC) faults [IEE06]. The OC faults mean there is no current flow in the faulty area, and they are normally caused by broken conductors [SKS+00, RBM01]. However, the majority of occurring faults are SC faults in the networks. A significant feature of a SC fault is the appearance of high current in the network, due to the negligible impedance and constant voltage. Thus, if a SC fault is not detected and isolated quickly, the disturbance could damage the electrical equipment with fires or explosions, or even large-scale blackouts. In this thesis, the research will focus on the protection against SC faults.

In three-phase AC power networks, SC faults are classified as: Single Line-to-Ground (SL-G) fault, Line-to-Line (L-L) fault, Double Line-to-Ground (DL-G) fault, and Phase-to-Phase-to-Phase (Ph-Ph-Ph) fault [EH08, Y. 13]. The Figure 2.7 illustrates all mentioned fault types, including the fault impedance Z_f. When the Z_f equals to zero, the fault can also be called a bolted SC fault [GSO12].

Generally, the electricity breakdown resulting from SC faults are due to the failure of insulation, the reason could be an overvoltage (e.g. a lightning strike on the circuit), insulation deterioration (e.g. physical crack or chemical pollution on the insulators), or inclement weather and environment (e.g. the wind blows down poles) [Y. 13]. A SC fault can take place on overhead lines or underground

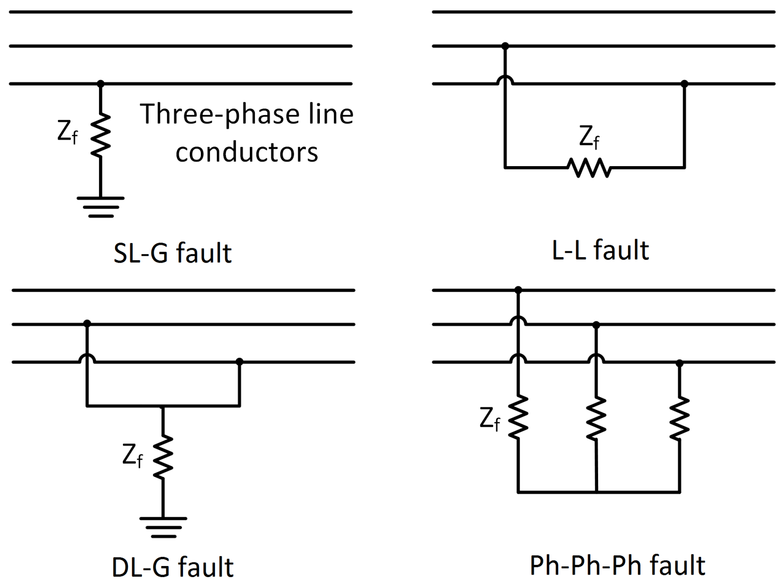


Figure 2.7: SC fault types in three-phase AC networks [Hon15]

cables. Table 2.1 lists the statistical proportion of the faults attributed to various power system components, where the fault statistics were gathered from 200kV to 250kV power transmission lines during the year of 2008 in the Poland.

It is apparent that the overhead lines and underground cables risk over half of the power systems' faults, and about 75% of them are registered as SC faults [EH95]. Therefore, in order to minimise the disconnection on overhead lines, thus ensuring maximum security of supply to consumers, many protection schemes have been applied in the power system protection, and the technologies utilised in protection devices are refurbished often to improve the system reliability.

2.4.2 Protection Devices

With regard to the characteristics of electrical faults, the distribution system could encounter different types of faults: OC or SC faults, internal or external faults, temporary or permanent faults. For the purpose of achieving a safe and reliable electricity service environment, various protection schemes with technologies are designed and applied into power networks. The most common schemes used for fault protection contain relays coupled with circuit breakers, sectionalisers, fuses and auto-reclosers. This subsection summarises the definitions of

Power system component	Prevalence of faults (%)
Overhead lines	50
Underground cables	9
Transformers	10
Generators	7
Switchgear	12
Other equipment	12
Total	100

Table 2.1: Proportion of the faults attributed to various power system components [Y. 13]

first three protection devices. The background and technology of auto-recloser will take the installed SPEN's PMARs as an example in the following subsection 2.4.3.

In general, a protection relay with a circuit breaker is used in the primary distribution systems for detecting the fault and isolating it upon the control and protection requirements. Figure 2.8 shows the common components contained in a typical protection relay system. The voltage transformer (VT) and current transformer (CT) are the measurement devices that sense the values of voltage and current from the power system overhead lines, and convert them into appropriate levels for the protection relay to deal with. The relay is the main element of the protection system which utilises integrated algorithms to determine whether to react to the detected fault or not. If the fault condition exceeds the threshold for action, the relay will register a tripping signal to the associated circuit breaker, and it isolates the fault immediately. For some protection technologies, e.g. differential protection, the measurement for detecting faulty condition is required with the support from additional information sources (e.g. the remote end protection relay on a transmission line) through communication connections. And sometimes, the communication is capable of sending the signal for tripping [Bla13].

From the years of manufacturing simple relays to meet the objectives of pro-

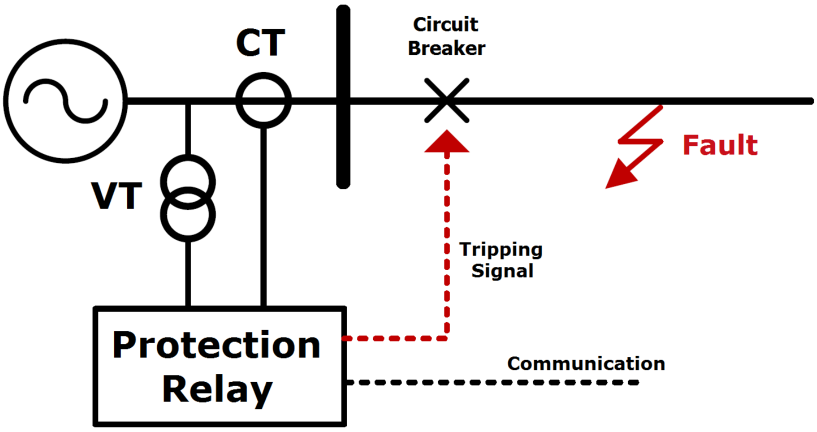


Figure 2.8: A typical protection relay system

tection in history, the relay technology has been developed over 100 years. As shown in Figure 2.9, the history of development can be split into four stages, from the earliest electromechanical relays to the state-of-the-art numerical relays. Electromechanical relays are constructed with electric and magnetic components (e.g. operating coil and contacts). They are the first type of protective relays, and rarely used today, because of the substantial amount of maintenance required to keep the moving parts operational. In the middle of the 1960s, static relays were introduced to the protection system with the significance of being faster and having a more accurate operation. These are based on semiconductor technology. However, the static relays could be unsuccessful in adverse environments due to the failure of one component in them. Until the 1980s, the digital relays with processing function improved the reliability in the protection systems. And around 1985, the next generation of digital relays - numerical relays, became an ever more common choice of relays when designing the protection systems [SM11b, Lun11, AB14].

Numerical relays are integrated with a programmable microprocessor, which can monitor real-time circuit conditions for processing and executing protection. The powerful microprocessor allows multiple protection functions simultaneously based on advanced computing technologies and large memory capacity. Meanwhile, communication function and digital signal converter of numerical relays

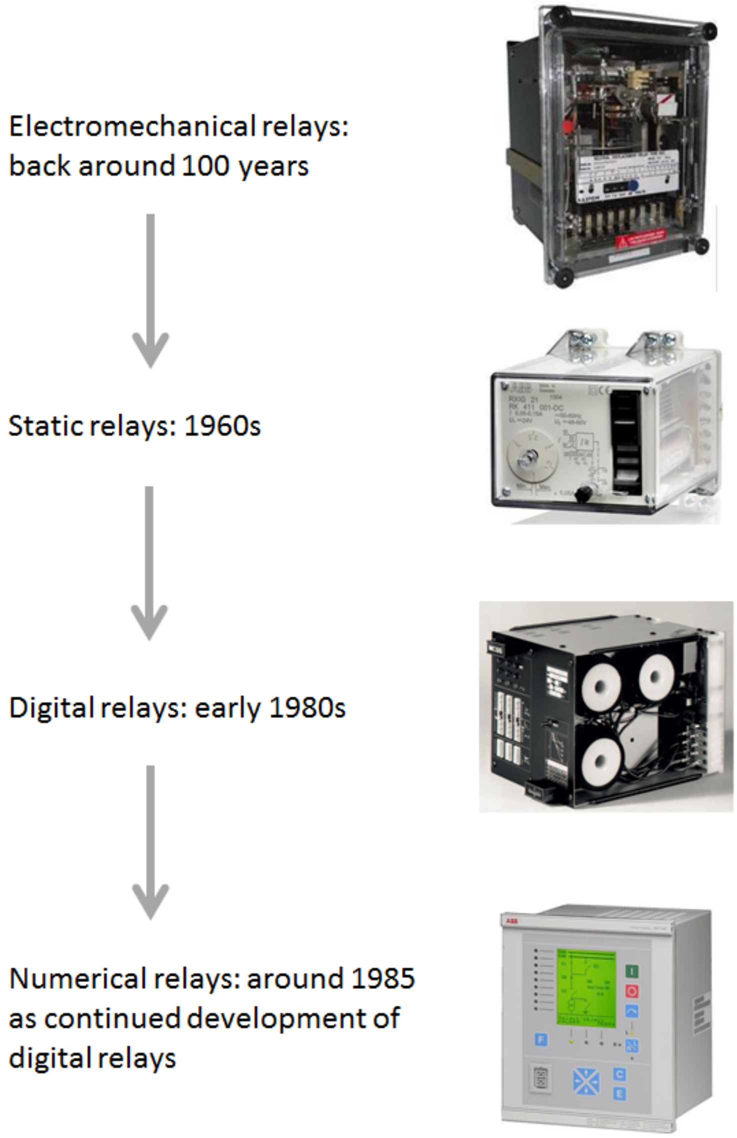


Figure 2.9: The development of relay technology [Gri11]

make the engineers' control more flexible. Numerical relays are referred as one type of protection IEDs [Gri11].

The sectionaliser is used for automatically isolating the faulted segment of distribution lines with other protective devices, such as reclosers. The sectionaliser only senses the current above a predetermined level, without considering the registered time period of overcurrent. And it counts the number of de-energisations with a recloser. If the number exceeds a preset value (normally set to 3), and the current of abnormal condition is still higher than the actuating value, the sectionaliser will open the circuit within its protection zone. Sometimes, the fault could be isolated by the recloser before the sectionaliser isolates the faulted area, so the sectionaliser would reset the counted number to zero for the future detection. [SM11b, Mom07, EAT14].

A fuse is typically used in radial distribution networks. The fuse is a single-phase device which operates by melting the metallic element to interrupt an overcurrent fault, and the fuses are not be reusable. Unlike a sectionaliser, fuses sense both the magnitude and duration of the fault current flowing through them. When the continuous current exceeds the threshold for respond, the process of interruption consists of two steps: thermal process and interruption process. At the stage of thermal process, the heat generated by the fault current is higher than that of normal condition, the high temperature will then lead to the melting of element. But after melting, the fault current would still flow through an arc, which takes some time to quench. Finally, the fuse interrupts the fault on the circuit, and the time period of the second stage interruption process is called the "arcing time" [Feh16, SM11b, Cla11, Das12].

2.4.3 Pole Mounted Auto-Recloser

The majority of devices installed in the distribution protection systems are auto-reclosers, rather than relays, sectionalisers, or fuses described in the previous subsection. They are used for isolating and mitigating the affected area when a fault occurs on the circuit. Typically, most overhead line faults in distribution networks are transient, only lasting a few cycles, such as a flashover on the cracked insulator [GCF16]. The causes of these faults are probably related to weather conditions or animals. If these faults do not clear within a prescribed time, they will result in the tripping of circuit breakers for isolating the faulted area, and disappear after the closing of circuit breakers. For the special purpose of avoiding unnecessarily long outages resulting from transient faults, the auto-reclosers are employed in the distribution networks. Depending on the characteristics of network topology and geography, overhead lines occupy most of distribution lines in the UK networks compared to underground cables, therefore, the Pole Mounted Auto-Reclosers (PMARs) are widely applied. Meanwhile, with consideration of European and UK's power networks' requirement, the PMARs are designed with single tanks for balanced three-phases. Figure 2.10 illustrates one type of PMAR installed on the overhead lines of SPEN distribution networks, including the main components for protection operations [SP 12b, ABB07, NOJ16].

As shown in Figure 2.10, a self-controlled PMAR contains two main parts: main tank is used for executing protection operation through the interrupter (when a fault occurs, the magnetic actuator will power the pushrod to draw out a vacuum room, which disconnects the electricity between two terminals; and when the pushrod resets, the vacuum room disappear and power will be restored) and the control cabinet is responsible for logic algorithm processing and communication. Generally, based on the interrupter's categories, the PMARs can be divided into oil interrupted and vacuum interrupted. For the control cabinet, various modules achieve different functions. The driver module controls the status of the interrupter automatically in the main tank, and the Main Processor Module (MPM) analyses the monitored circuit's conditions for reaction, and records details of fault events and anomalous activities within its memory as well. The integrated antenna in the control cabinet will communicate with RTUs for updating the tripping information to the central database, and receiving remote commands for operations. With these features, PMARs can be classified as protection IEDs in the distribution networks.

Furthermore, the PMAR's protection scheme is time-current coordination. The current sensor first detects the overcurrent anomalous activity, and registers

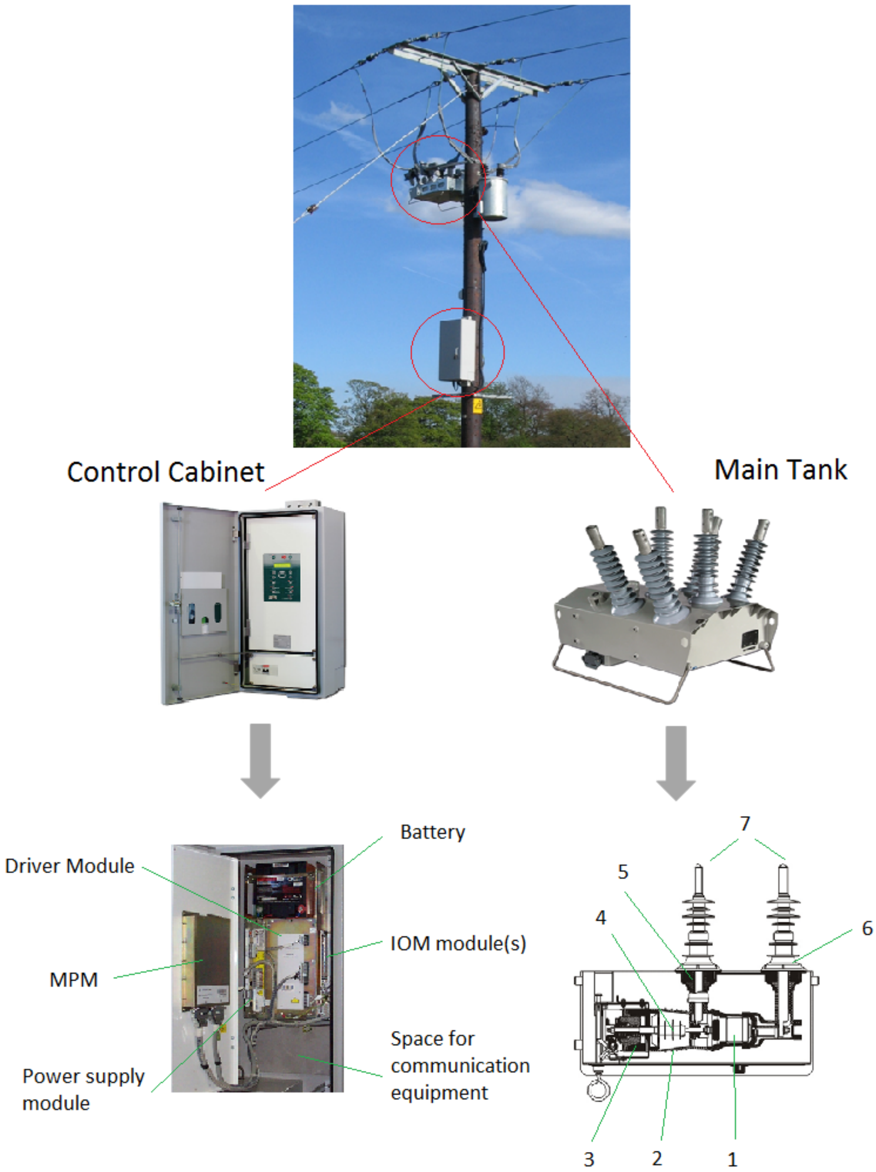


Figure 2.10: A typical PMAR with its main tank and control cabinet. 1 = Vacuum interrupter; 2 = Polycarbonate housing; 3 = Magnetic actuator; 4 = Pushrod; 5 = Current and voltage sensor; 6 = Bushing; 7 = Terminals [NOJ16]

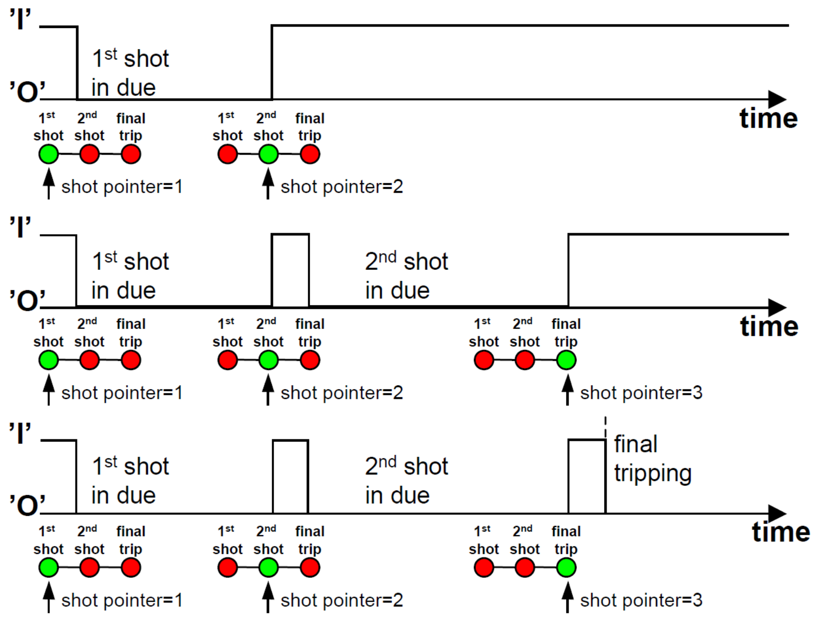


Figure 2.11: A typical sequence of PMAR's operation mechanism. Top: the 1^{st} AR-shot is successful; middle: the 2^{nd} AR-shot is successful; bottom: both AR-shots fail and lockout. 'I' = circuit breaker closed, 'O' = circuit breaker open [ABB07]

it in the MPM. Secondly, the MPM starts to monitor the time duration, if the activity lasts longer than the predetermined threshold for tripping, the MPM will upgrade the activity to a fault event and signal the driver module to command the tank for response. After a preprogrammed time, the contact will be reclosed automatically for restoration. The PMAR would try a preset number of attempts of open and close for eliminating the on-line faults. If the fault does not disappear during the cycles of the auto-recloser's operation, the PMAR provides the final tripping (i.e. lockout) to remain open, and it can only be reclosed manually or controlled remotely. Otherwise, the fault can be cleared and supply can be restored automatically by the PMARs. Figure 2.11 demonstrates the sequence of PMAR's operation mechanism [ABB07, Mom07].

Within the sequence of PMAR's protection operation, the automatic reclosing follows each tripping after a delay time (typically 10 seconds) is called an AR-shot, and the shot pointer counts the number of trip-reclose cycles. If the fault disappears within a present number of allowable AR-shots, the shot pointer will be reset to zero. Otherwise, if the fault still persists, and the counted number of the shot pointer exceeds preset value, the PMAR then actuate a final trip (i.e. lockout) to isolate the fault. In this figure, it shows the particular PMAR's sequence with SPEN's specific requirement, that is the shot pointer is set by 3 and the delay time for automatic reclosing is 10 seconds.

As described previously, the PMARs register the detected fault with overcurrent on the circuits in their MPMs, and the MPM determines the reaction decision based on the monitored time duration. However, the transient fault could clear itself without the PMAR's operation, and this activity will be also recorded in automatically produced log files. In general, the PMAR log file contains data that can be divided into three main parts [WSKM14] (examples of the original PMAR log data were shown in Appendix A to demonstrate the three categories of data information respectively, and to display the data format and structure of the log file):

- Fault event, which leads to trips or lockout operation of the PMAR. The MPM will record the time stamp of the occurrence and clearance of this fault, each time stamp of responded tripping or lockout, current amplitude of fault, and affected phases, etc.

- Abnormal activity, which does not lead to a trip operation. The MPM will store the time stamp of registration and disappearance of the activity, current amplitude, and affected phases, etc.

- Device event, which reflects the conditions of each module integrated in the PMAR's control cabinet. The MPM will produce the alarm messages related to the problematic modules in the log file.

Moreover, the PMAR log file must contain the basic circuit information, such as specific PMAR's name and circuit code (concerning the privacy of DNO data, the basic circuit information was hidden in the examples in Appendix A), which corresponds to the SCADA alarm database. This thesis will analyse all the above data for fault diagnosis and prognosis to minimise the supply interruptions and improve the system performance.

2.5 Performance of Distribution Systems

Distribution systems are the secondary electricity transmission networks which deliver the power to the consumers directly. Any temporary interruptions or permanent outages within a distribution system will affect the customers adversely. Therefore, the performance of distribution systems becomes one factor of evaluation criteria of penalty and reward for the distribution network operators, which covers the areas of reliability, availability and customer satisfaction. This section will introduce the performance on the circuits with effects of different fault scenarios, then briefly describe the reliability of distribution systems and how to assess the performance through the reliability indices [CK09, Bau10, BA13].

2.5.1 Types of Faults

The previous subsection 2.4.1 classifies the electrical faults into OC and SC faults based on the current flow characteristic. In addition, the overhead line faults can also be divided into supplementary categories depending on their performance and the operation of deployed PMARs. Generally, the three types of faults are transient faults, semi-permanent faults and permanent faults [WSMK15, NGW07]:

- Transient fault, as stated in the background review of PMARs, is an inevitable temporary fault driven by the external environment (i.e. weather, animals, etc.) which occurs in the networks and might lead to a short-term supply interruption. According to whether the fault causes a PMAR operation or not, the transient fault can be sub-divided into two following classes with typical examples [ABB07]:

  (a) Self-clearing fault, which represents the fault clearing itself without an operation from PMARs. For example, it could be a developing arc with small value of current and voltage. Although probably the MPM of a PMAR could detect the arc, it can be extinguished in free air before allowing the protective device to take action.

(b) Non-damage fault, is characterised by there being successful operations (i.e. AR-shots) of PMARs but which does not result in a long-term power outage in the protection zone, due to the fast restoration scheme of the PMARs. For instance, a falling tree branch touches the three-phase uninsulated overhead line and causes a SC fault which leads to a tripping operation by a PMAR. After the preset delay time, the circuit is automatically reclosed for restoring supply without a lockout or a need for permanent repairs.

- Semi-permanent fault, is a class of evolving fault (defined by SPEN experts when categorising the overhead line behaviour) arising from the degradation of overhead lines which leads to frequent short-term supply interruptions. The semi-permanent fault could be either isolated by successful AR-shots like a non-damage fault, or isolated by a lockout of the PMAR following failure of AR-shots. An example is rain affecting a cracked insulator on a wood pole resulting in trips or lockouts of the PMAR (which can dry out and then no longer provide a fault path).

- Permanent (or damage) fault, is a prolonged power outage which is isolated by the operation of protective devices, such as the lockout of a PMAR. This type of fault only be isolated remotely with a control command and be repaired manually with delivery of maintenance staff. For example, the storm blew the trees down on the overhead lines, which disrupted the electricity supply, and should be recovered by the maintenance staff.

With the view of three different fault scenarios, the semi-permanent fault is an intractable problem due to the difficulty of identification. Not only does the semi-permanent fault affect the quality of daily electricity service (e.g. resulting in tripping event due to frequent transient fault activity), but it may evolve into a more serious permanent fault resulting in long-term outages. Meanwhile, it is evident that the PMAR is a solution that allows comprehensive protection for the overhead line. If it is correctly designed and implemented to match the local regulations and satisfy the protection requirement, the PMAR will provide a significant and increasing benefit in defending against the interruptions caused by transient and semi-permanent faults. This will enhance the reliability and potentially improve the performance of distribution systems.

2.5.2 Performance with Reliability Indices

The definition of performance of distribution systems varies slightly across different countries due to the network regulations, so the key indicators against the performance are not always the same. However, there is one factor referred to for evaluating the performance around the world, which is reliability. It is the ability of keeping a predetermined satisfactory level of quality and security when delivering the electricity to the customers. Utilities and regulators could assess the performance through reliability analysis. To quantify the reliability of distribution networks, a set of reliability indices are defined to measure and recognise the performance [Sta09, IEE12]. In this thesis, the commonly accepted indices defined by Institute of Electrical and Electronics Engineers (IEEE) and the similar indices developed and used extensively throughout the UK's industries will be both introduced. These applied indices focus on indicators for the annual average performance of distribution networks, in terms of frequency and duration of supply interruptions. They will consider the performance based on the number of customers whose supplies are affected. The details of basic reliability indices are shown as the following equations [Mom07, NGW07]:

1. SAIFI (System Average Interruption Frequency Index) is defined as:

$$SAIFI = \frac{\text{number of customer interrupted}}{\text{total number of customers served}} \text{ during the period}$$

$$SAIFI = \frac{\sum_{i=1}^{R} N_i}{N_T} \tag{2.1}$$

where

R = number of sustained interruption event during the reporting period

N_i = number of customers interrupted by ith interruption event

N_T = total number of customers served in the assessed area.

SAIFI describes the average number of interruptions experienced per customer over a period of time in a particular area. If the total number of customers in the area is fixed, the only way to reduce SAIFI and improve the performance will be to mitigate the supply interruptions in the distribution systems.

2. SAIDI (System Average Interruption Duration Index) is defined as:

$$SAIDI = \frac{\text{sum durations of customer interruptions}}{\text{total number of customers served}} \text{ during the period}$$

$$SAIDI = \frac{\sum_{i=1}^{R} U_i N_i}{N_T} \tag{2.2}$$

where

R = number of sustained interruption event during the reporting period

N_i = number of customers interrupted by ith interruption event

U_i = restoration time after ith interruption event

N_T = total number of customers served in the assessed area.

SAIDI determines the average duration of interruptions distributed on each customer in the particular area. The known SAIDI plainly reflects the time duration of customers without electricity service. The unit of SAIDI is usually in minutes or hours.

SAIFI and SAIDI both are main universal indexes for evaluating system reliability. However, in order to achieve specific business goals and meet regulations within organisations, some industries and utilities have their own measures to appraise the reliability. For example, the distribution network operators in the UK use self-defined reliability indices instead of SAIFI and SAIDI and report to the regulators [SP 13, Bri11, ofg09, ofg05, ofg02]. The same values are:

1. CI (Customer Interruption) is defined as:

$$CI = \frac{\text{number of customer interrupted per 100 customers}}{\text{total number of customers served}}$$

$$CI = \frac{\sum_{i=1}^{R} I_i}{N_T} \tag{2.3}$$

where

R = number of sustained interruption event during a year

I_i = number of customers interrupted per 100 customers by ith interruption event

N_T = total number of customers served in the assessed area.

From the apparent view of mathematical equation, CI is the same index as SAIFI, indicates the average customer experiences a sustained interruption in every 100 customers over a period of time (usually annually) over all incidents, where the supply interruptions should last more than three minutes, excluding the re-interruptions caused by the same incident.

2. CML (Customer Minute Lost) is defined as:

$$CML = \frac{\text{sum lost minutes of customer interruptions}}{\text{total number of customers served}}$$

$$CML = \frac{\sum_{i=1}^{R} U_i N_i}{N_T} \tag{2.4}$$

where

$$R = \text{number of sustained interruption event during a year}$$
$$N_i = \text{number of customers interrupted by } i\text{th interruption event}$$
$$U_i = \text{customer minutes lost for } i\text{th restoration}$$
$$N_T = \text{total number of customers served in the assessed area.}$$

Like the SAIDI, CML displays the average power outage time of per customer per year ($mins \cdot yr^{-1}$). The CML only considers the minutes lost where the interruption lasts for three minutes or longer.

For SAIFI and SAIDI or CI and CML, they all indicate the quality of customers' electricity service and the performance of distribution networks. A network kept in good condition will have fewer supply interruptions. And a network which is affected by frequent and lengthy faults will lead to poorer reliability indices. Therefore, the major target of improving the performance is to reduce the CML and CI of distribution systems in the UK.

From the above discussions in Section 2.4 and 2.5, it can be concluded that the challenge is to improve the performance of distribution networks, which is not simple. Hence, many utilities apply the auto-reclosing protection scheme into their systems to assist the reliability improvement, and this protects the networks against different on-line fault scenarios (i.e. transient faults, semi-permanent faults and permanent faults) actively. However, the employed auto-reclosers could only respond to existing faults and not prevent them from occurring. In order to mitigate the supply interruptions and decrease CML and CI from the root cause, the large volumes of undiagnosed data generated from the auto-reclosers and associated SCADA alarm data should be deeply analysed. This is the main motivation for the research presented in Chapter 4, where a designed and developed decision support system is proposed to address these challenges.

2.6 Existing Research Activities and Commercial Systems Associated with Fault Diagnosis and Prognosis in Distribution Automation

This section provides a review of research activities and commercial systems that are relevant to fault diagnosis and prognosis for distribution automation, which are mainly focused on utilising the SCADA alarm data or protection IED (e.g. PMAR) data for analysis. Finally, a discussion of the shortcomings of these techniques and the potential improvements for filling the gaps are provided.

2.6.1 Systems with SCADA alarm processing

As stated previously, the SCADA technique has been widely deployed in the power electrical systems. Consequently, many in the research community have concerned the large amount of SCADA alarm data for fault diagnosis and prognosis to improve the system control and protection. The work covers the areas of generation system (e.g. the analysis of SCADA data associated with wind turbine operation to detect the equipment's degradation [QFS+16]), the transmission system (e.g. parameter estimation on the transmission lines by using SCADA data [MSAA15]), and distribution system (e.g. SCADA data supports distribution feeder models to predict the states and measurements at the buses [HLL15]). Research of SCADA alarm data in DA fault diagnosis and prognosis can be categorised into two main aspects: one only concentrates on processing SCADA alarms for distribution fault diagnosis, the other one enhances systems of fault analysis by combining the SCADA alarms with available IEDs' data or network data. A brief review of these approaches is provided as follows.

In [TLZ+14], an alarm management framework was designed and developed to help recognise event sequences from SCADA alarm streams and to accelerate engineers' decision making. It automatically categorised the effective events for the control centre and reduced redundant, even dubious, alarms by applying complex event processing technology into the framework. In the proposed alarm processing framework, the defined constraints combined with complex alarm processing technology to recognise isolated events and alarms into integrative events. That is to say, when the distribution system is affected by the occurring events, the associated alarms will be first detected and classified into different events (e.g. fault events (reflected by a sequential relay actions or circuit breaker status changes), reclose events (i.e. AR-shots), or self-healing events (manual reconfigurations), etc.), the effective events could clearly represent the current conditions of circuit and assist decision making. However, this framework only detects and integrates the fault events from the SCADA alarms and provides them to the control centre. Further fault diagnosis for decision making is not included.

Compared to the previous research work, [CBG14] introduces a method to evaluate the fault diagnosability from the provided SCADA alarms for assisting operators' diagnostic process. In the lists of SCADA alarms, some alarms are raised based on the occurrence of different fault conditions, while others may be specific to a particular fault which is valuable for diagnosis. This proposed method will discriminate the diagnosable fault from a set of other faults by calculating the index of relevance of each alarm. The relevance is determined by how often the alarm is (or seldom) raised by a particular fault, and seldom (or often) it is raised by the other faults. A high value of index indicates the fault is associated with the relevant alarm, which should be focused on diagnosis. This methodology offers meaningful alarms for fault analysis by filtering the inactionable alarms. This mimics a pre-processing stage for fault diagnosis to accelerate decision making.

Apart from these systems for assisting fault diagnosis, [Che12] presents a decentralised fault diagnosis system for fault section estimation, which details the hardware implementation of field-programmable gate arrays. This system obtains the SCADA alarm data directly from RTUs instead of the control centre to avoid the influence of communication problems. The graphic models are adopted for knowledge representation in the proposed fault diagnosis algorithm, the nodes and arrows represent performance (i.e. circuit breaker status change) and related conditions (e.g. fault occurrence), and implicational rules describe the relations between cause and effect. These relations will provide information support for fault diagnosis. Therefore, based on the detection of a SCADA alarm message, the fault diagnosis system can estimate the possible fault section with the provided operation information. Additionally, it can be integrated with existing SCADA systems for on-line analysis as well.

Other than the hardware implementations of fault diagnosis with SCADA alarms, the early research work looked at intelligent systems (software implementation) to provide decision support assistance by analysing SCADA alarm data for fault diagnosis, which were developed by Burt *et al.* [BMK+95], Vale *et al.* [VF98] and Burrel *et al.* [BI98] in 1990s. After these, [BMM+98] and [SLK+00] started to interpret Digital Fault Recorder (DFR) data for fault classification and protection validation, but these existing systems only interpreted the DFR data for fault diagnosis. However, McArthur *et al.* [MBM+98] demonstrated the integration of SCADA alarms and DFR data could enhance data interpretation with protection validation to support diagnostic systems. And in 2002, Hossack *et al.* [HMMM03] developed a flexible and scalable open architecture by using Multi-Agent System (MAS) technology with automatically retrieving and interpreting SCADA alarms and DFR data for providing disturbance diagnosis assistance to protection engineers, which was known as Protection Engineering Diagnostic Agents (PEDA) and applied into SPEN's distribution networks for on-line analysis.

In general, PEDA is an automatic diagnosis system replacing the manual process of analysing post-disturbance. For the traditional manual approach to the power system diagnosis, the SCADA alarms occurring near the time of a particular disturbance should be firstly gathered and selected to identify the disturbance and related events. Then, through the identification of location and nature (e.g. transient SC fault) of the disturbance incident, the protection engineers can determine the additional data sources (e.g. DFRs) to retrieve and to interpret the useful disturbance information for fault diagnosis. While, this is a time-consuming and problematic task. The PEDA designed separate analysis tools (e.g. multi agents) to interact and cooperate. These agents can analyse the data individually and share it to produce a clearer analysis view, from the collection and interpretation of the suitable data to final disturbance diagnosis. In this research work, the agents were developed to achieve the functionalities of gathering SCADA disturbance incidents and interpreting the additional data sources.

McArthur *et al.* conducted extensive research on PEDA for further protection validation and disturbance diagnosis to help assess protection system performance [MD06, MDC+07a, MDC+07b, NDMM09]. The extended system comprehensively analysed SCADA alarms, DFR data, protection settings and assets databases. The entire automated process not only contained the previous designed functionalities, but added protection validation and a diagnosis agent. The model-based reasoning engine integrated with the agent support diagnosis by comparing the simulated protection behaviour with actual operation observed by DFRs. Meanwhile, the upgraded number of engineering assistant agents in the PEDA could inform protection engineers with new diagnostic information when it becomes available. [DMM+06] presented the multi-agent system technology in the PEDA and demonstrated the robustness and flexibility required for the on-line post-diagnosis in SPEN distribution network. With the perspective of the development of MAS technology deployed in disturbance diagnosis, many researchers conducted diverse work focusing on the fault diagnosis by utilising SCADA alarms and available network data, such as [EDM+13] designed an automatic model-based diagnosis system for protection performance assessment and incident identification with SCADA alarm processing. Moreover, [RKD+11] proposed a rule-based system to monitor the performance and health of distribution automation through analysis of SCADA alarms.

The above work on automatic disturbance diagnosis for assisting protection validation and performance evaluation is very relevant. This is because the SCADA systems are generally deployed in modern power distribution networks for protection and control. And SCADA and IEDs (e.g. DFRs) data contain abundant information reflecting the circuit and equipment conditions, which could be retrieved and interpreted for diagnosis. However, many in the research community have not yet analysed the SCADA alarms or IED data for fault prognosis yet. Additionally, protection IED (e.g. PMAR and DPR) data is rarely taken for disturbance diagnosis while the data also records the details of disturbance and protection action.

2.6.2 Activities with protection IED data analysis

As discussed in the subsection 2.6.1, the analysis of DFR data integrated with SCADA alarms could assist fault diagnosis in distribution automation. This is because the high-frequency captured by the DFR data contains various information for automated event analysis. Compared to the DFR data, modern protection IED (e.g. DPR or PMAR) data not only describe the circuit performance (e.g. manual protection operation or AR-shots), they also has the capability to record event details. This includes fault location, current and voltage amplitude of fault, affected phases, and so on. Therefore, research on analysis of protection IED data for fault diagnosis and prognosis has been increasingly conducted in recent years [KPS+10, ZDGK10, KG09]. In this subsection, several examples of work related protection IED data for fault diagnosis and prognosis are presented and reviewed.

'REZAP Fault Master' is one type of PMAR installed in Holland's distribution network to reduce the number of repeat interruptions causing decreases in CML and CI. [DZG13] proposes a method to identify the fault location based on analysis of the records and events stored in the REZAP server. In the presented method, the fault location is identified based on the calculation of fault impedance through a fault location algorithm called Single Ended Location of Fault (SELF), in which the REZAP data provides the voltage and current waveforms to calculate the fault impedance. Like REZAP, 'BIDOYNG' is another kind of auto-recloser which is employed by one of the UK's distribution network operators. It has the ability of storing and communicating the detected anomalous events combined with protection scheme, which can be referred to as a protection IED. [GR13] demonstrates the fault location detection by utilising the SELF algorithm to calculate the fault impedance based on the supported voltage and current waveforms. And in this research, a pre-fault voltage disturbance is apparent from the observation of the fault waveforms. The authors exhibit the characteristics of these disturbances and suggest the detection of pre-fault disturbance could mitigate the supply interruptions before the fault occurs in future work.

Apart from fault location diagnosis through analysis of protection IED data, [XK09] presents a complete system solution for automated integration of IED data captured from distribution networks to generate diagnostic report for protection engineers. In this developed system, the DFR data, DPR data, and circuit breaker monitored data will be collected and processed for real-time analysis. Based on the information exchange between different types of IED data, the automated application provides the protection validation and disturbance diagnosis of relay operations by using logic and cause-effect chain, which compare the expected relay behaviour with the actual one. If the validation result is not consistent with the performance, the event report will be automatically generated for protection engineers. Meanwhile, some concise advised actions should be taken immediately by engineers, which are suggested in the event report through the intelligent analysis of checking the settings and predefined logic rules.

Kezunovic *et al.* [Kez11, PKK15, GK09] extended the previous research into several different applications: optimal fault location, alarms processing for disturbance diagnosis, and fault detection, and classification with analysis of SCADA data and IED data separately or cooperatively. Expect for the discussed fault location identification and alarm processing for protection assistance, the application of detecting and identifying the fault types is a major support to fault diagnosis as well. In the designed system, the application applies a neural network to train the collected DPR data (i.e. the input voltage and current waveform signals), and the fault classification is processed through the K-Nearest Neighbour (KNN) algorithm to cluster fault patterns. These identified fault patterns could be described by engineers' knowledge and experience and translated into rules for helping decision making when the similar faults are detected in future.

2.6.3 Existing Systems that May Be Used for Fault Diagnosis and Prognosis in DA

With respect to the review of analysis of SCADA alarm data and protection IED data for fault diagnosis and prognosis in DA, it is evident from there are existing commercial systems or research activities that may be developed and used for further fault diagnosis or prognosis, although currently they can not be used for this purpose. For example, [MBM+98] could add different agents for protection operation prediction with analysis of DFR data, or [GR13] could predict a fault event by detecting the pre-fault disturbance and [Kez11] could utilise the clustered patterns for fault prognosis with matching the suitable DPR data. The common characteristic of such systems is that they all provide a platform that allows the processing of information hidden in the provided data for additional functionalities. The shortcomings of such systems for fault diagnosis and prognosis in DA are:

- The systems or activities mainly focus on validating the protection operation on the circuits, and generating diagnostic reports for assistance, which do not give the reasons of the disturbances or unexpected operations.
- The systems usually diagnose the disturbance based on checking with predetermined settings of the protection system; rarely considering protection engineers' knowledge and experience. These can be used for fault diagnosis
- The systems analyse the SCADA alarms and IED data but provide no fault prognosis function.
- All the systems which take the protection IED data into analysis, focus only on part of protection operation data (e.g. circuit breaker status change). Although the protection IED data used in the research [Kez11, PKK15] also could provide more information of anomalous activities, these undiagnosed details cannot fully support the analysis of identifying the root causes of fault events. However, the PMAR data (detailed in subsection 2.4.3) can

be used for fault diagnosis and prognosis to assist protection and asset management.

2.7 Conclusion

This chapter has provided a review of the fundamentals of distribution systems with a focus on the topics associated with protection and automation. With the growing complexity of distribution systems' design and implementation and increasing development of DA technologies, it can be concluded that the risk of electrical faults is inevitable for any power systems. Meanwhile, the data captured from transient interruptions or permanent outages allows the reactive fault management and post fault investigation into anomalous network behaviour, which could support detection and prediction of evolving circuit's faults. To mitigate the effect of the faults, and quickly isolate the faulted area and restore power manually or automatically, the SCADA system and intelligent DA devices (e.g. PMARs) are deployed into the distribution networks to improve the system performance. Furthermore, the reliability indices can be evaluated to assess the performance, which are applied by the distribution network operator around the world.

The main research activities and commercial systems associated with fault diagnosis and prognosis in DA by utilising the SCADA alarm data and protection IED data have been shown and reviewed in this chapter alongside their fundamental shortcomings. That is, many of the developed systems do not provide fault prognosis functionalities, and the disturbance diagnosis were rarely supported with analysis of details contained in the protection IED. The shortcomings of these existing systems are addressed by the methodology and design approaches developed through the research in this thesis, and mainly reported in Chapter 4.

Chapter 3

Review of Artificial Intelligence Techniques for Fault Diagnosis in Distribution Systems

3.1 Overview

The origin of 'Artificial Intelligence' (AI) is attributed to Warren McCulloch and Walter Pins in 1943 [RN10], and the term was first coined in the 1950s, where they proposed a model of artificial neurons with the capability of computation where the designed networks of neurons could learn. The birth of AI can be marked at the workshop in Dartmouth College (US) in 1956. During this workshop, McCarthy *et al.* stated that a machine can be made to simulate learning and intelligence, and attempts were made to describe how a machine can use language and concepts to solve problems traditionally requiring humans [RN10, Fog06]. So, what is AI? Nowadays, the pressing question is still not easy to answer, even though the field is almost 60 years old. The definition of AI can vary along different perspectives, Figure 3.1 shows definitions concentrating on computers 'thinking' and 'acting' intelligently.

The primary concern of AI is to use computational techniques to make machines behave (i.e. think and act) as intelligently as human beings for problem-

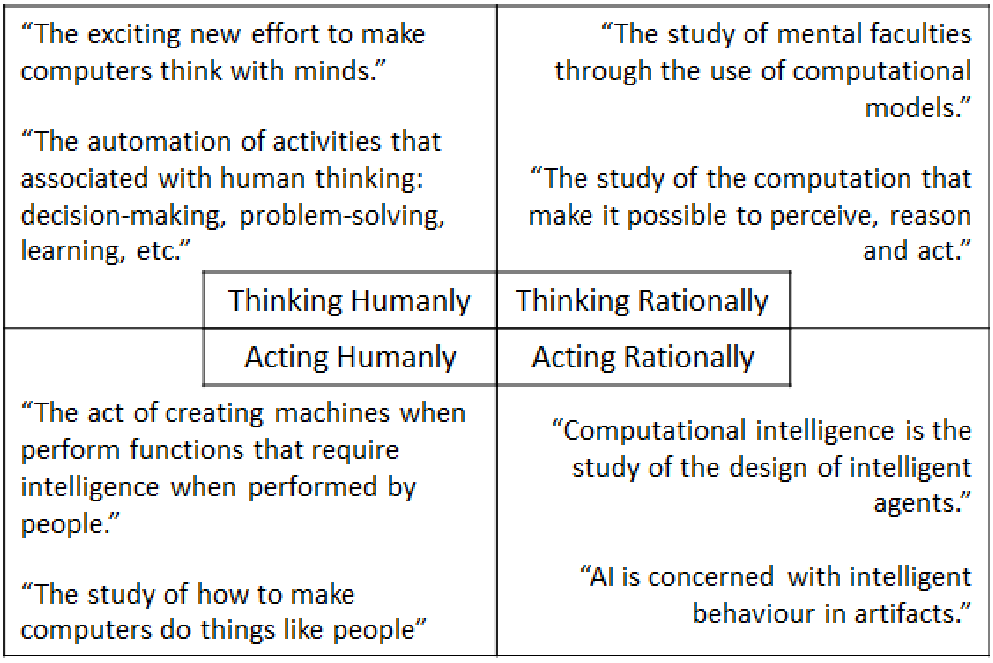


Figure 3.1: Some definitions of AI, classified into four categories [RN10]

solving, planning, reasoning, learning, and communication. Meanwhile, this thesis describes a Knowledge-Based (KB) system utilising the knowledge from experts to diagnose and prognose faults automatically. For this, there is an appropriate definition which is that AI is the branch of computer science which automates intelligent behaviour [Lug09].

Since the start of the study of AI, it has attracted plenty of research and discussion attentions covering lots of ground with successfully designed and developed applications, such as game playing, automated reasoning, and human behaviour modelling. These areas contain machine learning that improves machines' behaviours through learning from previous experiences. Two dominant approaches are: modelling human performance to simulate the process of solving problems like humans, and KB systems that solve problems using specific domain knowledge. Therefore, AI is relevant to any task that requires intelligence [Lug09]. Particularly, the commercial KB system deployed in the power industry improved the performance of utilities and customers in recent years. Such successful applications use domain knowledge to address problems for experts and maintenance staff in tasks such as; alarm processing, condition monitoring, fault diagnosis, and remedial actions for service restoration [M. 94].

In this chapter, a number of extensively-used AI techniques will be introduced, which are relevant to the problems described in the thesis. These are techniques usually used for diagnosis and prediction of equipment and system behaviour. These reviewed AI techniques will detail some associated applications for automatic fault diagnosis in the power systems. The remaining AI techniques which out of the scope of the research area reported in the thesis are not considered and related further information can be found in [RN10].

3.2 Artificial Neural Network

3.2.1 Overview

An Artificial Neural Network (ANN) is an AI technique that belongs to the field of machine learning dealing with the study of programming previous experiences with computational technologies automatically to improve their behaviours [Lug09]. ANN provides a general, practical, and robust approach for learning real-values, discrete-values and vector-values from examples, notably in interpreting complex sensor data captured from the real world [Mit97]. In this section, ANN and associated applications will be investigated.

The inspiration of ANN research comes from the observation of biological neural networks, which are built of vast numbers of neurons [Mit97]. The human brain is estimated to contain approximate 10^{11} neurons densely interconnected, and each neuron is connected to about 10^4 other neurons [Alp10]. The neurons are operating in parallel and the connections between them support the typical excited or inhibited neuron activities. As such, these parallel activities form the basis of human thinking and performance with surprisingly quick speed of complex computations [Mit97]. Therefore, the ANN system is aimed to model the complexities of a biological neural system for achieving highly parallel computation capabilities.

Based on the description of biological neural networks, ANNs are formed by interconnected processing elements referred to as artificial neurons. Each neuron

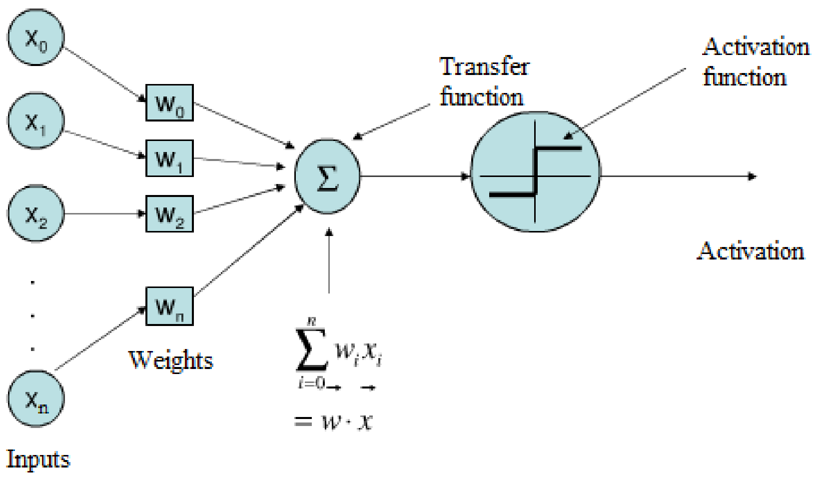


Figure 3.2: ANN processing element (artificial neuron) [Mit97]

has a number of inputs and provides a single output after computation. Figure 3.2 shows the composition of an artificial neuron with the ability of mathematical modelling, where x_i and w_i are the input and the weight of each input. Each input is multiplied by its corresponding weight, and then all weighted inputs are summed and transferred to determine the activation decision using the related activation function.

Different types of ANN systems use different numbers of artificial neurons based on the complexity of the learning task. Generally, an ANN contains one processing element called a perceptron [Mit97]. A perceptron calculates a linear combination of input values and uses threshold function to decide the output signal. That is to say, if the calculated value exceeds a pre-determined threshold, the neuron will output 1, otherwise -1. As a result, a single perceptron can be applied to represent many boolean functions, and the learning process of it is effectively providing the expected outputs by adjusting the weights related to input values [AS97]. However, for some complex learning tasks (e.g. interpreting visual scenes), ANN systems with multi-layers of interconnected artificial neurons are used instead of the single perceptron. These systems consist of an input and output layer visible to the real world, and intermediate hidden layers like a 'black box' of artificial neurons connecting to the input and output layer. Figure 3.3 illustrates a typical ANN, where there is 4 inputs, 5 neurons, and 1 input

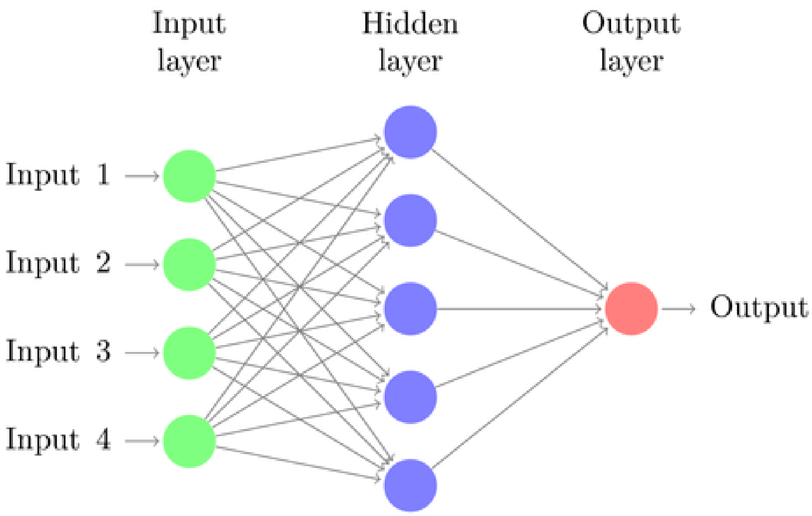


Figure 3.3: A typical ANN [Lug09]

respectively in the input layer, the hidden layer, and the output layer.

A number of algorithms presently exist in the learning process of ANNs. The back propagation algorithm, also known as the 'multilayer feed-forward' algorithm, is the commonly used method in ANNs, especially in pattern recognition tasks such as image classification and speech recognition [RN10]. The significance of the back propagation algorithm is to calculate the difference (i.e. errors) between the target and actual outputs, and propagate the errors backward to the forward layers to find the suitable set of weights which can minimise the errors through an effective gradient descent optimisation method. To achieve the best problem-solving, ANNs should have properly defined inputs and outputs. For example, the numerical values represent the feature describing the objects in the real world. Based on the characteristics of ANNs, they are mainly used in areas where sufficient data can be provided for processing such as the associated applications in forecasting sea levels [FTKM12], face recognition [Mit97], etc. In recent years, an advancing branch of ANNs is Deep Learning (DL) [BS14, GB10]. The DL neural networks decompose input data and learn features by themselves and have more than two hidden layers to address complex tasks rapidly, such as fast image processing [NVI16]), and fault diagnosis with partial discharge data in power system [CS15].

3.2.2 Applications of ANN Systems for Fault Diagnosis

The ANN technique has been extensively applied to the domain of power distribution systems, including fault location identification [JBRRGM14, CdVR07], improvement of protection performance [CJ98], fault cause identification [XC06], and reconfiguration for reducing power losses [SGR06], etc. Many applications integrate ANN techniques with other intelligent algorithms to improve the performance of research work.

The paper [TKV05] proposes an approach to estimate fault location by analysing available measurements from the substations or protection devices, which utilises the ANN technique to replace traditional fault section estimation methods. In the presented work, the ANN aims to classify Source Short Circuit (SSC) levels (i.e. the range of load conditions) associated with fault location when a fault occurs based on the monitored circuit's conditions. Prior to inputting the data for ANN learning, the captured data will be pre-processed to identify the fault types. Figure 3.4 shows the structure of developed ANN for fault section estimation. The inputs are pre-processed three-phase voltage and current amplitudes, and the single output is the line reactance with classified SSC level and fault type. In the structure of the ANN, the number of inputs, outputs, and layers are determined with experts' experience, and the training data is captured from 52 buses of 3 feeders in the distribution networks. Through testing the developed system, the observed results indicate the adopted ANN method provides more accurate fault locations compared to analytical methods.

Within the field of fault diagnosis in power systems, ANNs have demonstrated their efficiency in pattern recognition and forecasting. However, they usually require large volumes of training data (i.e. fault events) to improve accuracy in real-world scenarios which may be encountered in the field. Meanwhile, a potential challenge of ANNs is to ensure the network is not "over trained". This means when fault profiles are too precise, the misclassification of new faults will occur during the data training. Therefore, the successful and efficient ANN systems should prevent themselves from being over trained to increase the user confi-

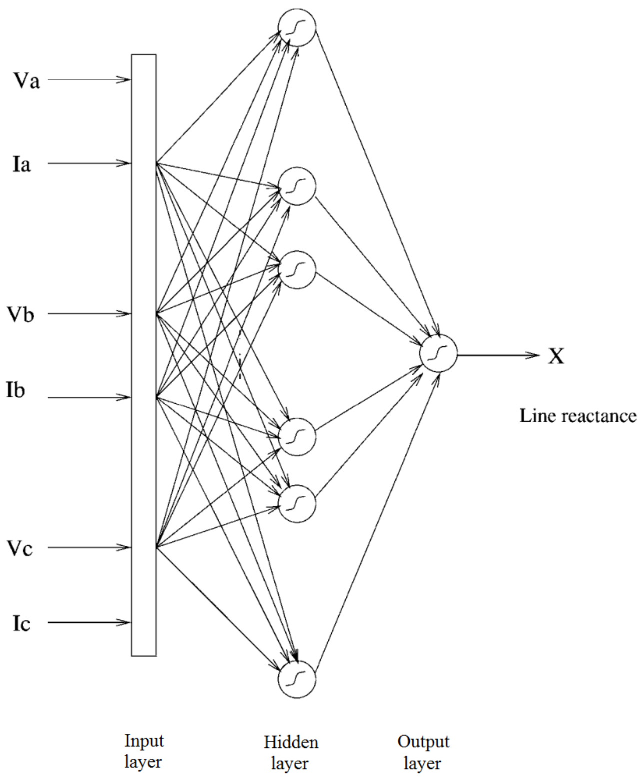


Figure 3.4: Typical ANN to determine line reactance in a fault section estimation application [TKV05]

dence.

3.3 K-Means Clustering

3.3.1 Overview

Clustering is an unsupervised learning method in the field of machine learning for problem-solving. It focuses on learning a mixture of parameters for unlabelled data and trying to organise them into different clusters [Alp10]. A cluster is a data set consisting of grouped vectors with similar characteristics. To be more specific, if a vector contains n features which are plotted as a point in an n-dimensional space, then a cluster is a region with a higher relative concentration of points compared to other regions. K-Means is one of the simplest clustering algorithms. It determines patterns and groups the data into K clusters [Str06, Rud10].

The main task of the K-Means algorithm is to position the k centroids (i.e. the centre of each cluster) and assign the points to the closest centroid. When all points are assigned to the closest centroids, the centroid of each cluster will be recomputed and updated based on the present points assigned to the cluster. Then, a loop for repeating assignment and updating centroids is generated, the algorithm will not finish until no point changes clusters, in other words, until centroids remain unchanged [TSK06]. Figure 3.5 shows the procedures of assignment and relocation of centroids using nine points and three clusters.

In order to obtain the optimal clusters representing the grouped points with similar characteristics, the initial placement of centroids is important, because different starting locations may lead to different end results. The appropriate choice is to set the k centroids as far as possible from each other. Meanwhile, to determine which cluster each point belongs to (i.e. the closest centroid to the point being assigned), typically, the Euclidean distances between the point and centroids are measured. To obtain accurate clusters, centroids are recalculated with the repeat algorithmic iteration, the squared error e_k^2 is quoted, which sums the squares of the Euclidean distance measured between each point and the

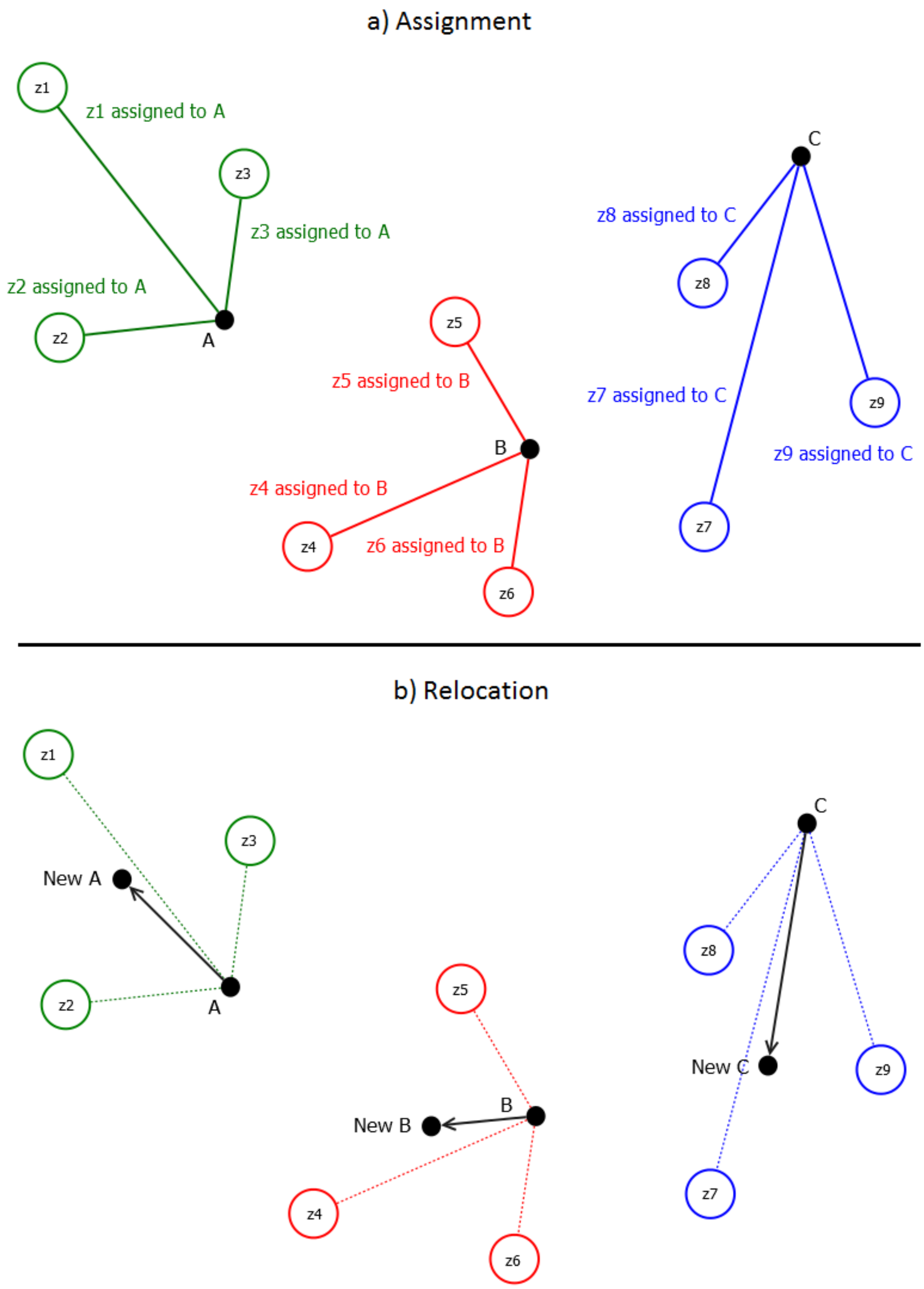


Figure 3.5: Main steps of K-Means algorithm processing

assigned centroid, and defined in the following equation:

$$e_k^2 = \sum_{i=1}^{n_k} \left\| x_i^{(k)} - c_k \right\|^2 \tag{3.1}$$

where

$$n_k = \text{number of points in the } k^{th} \text{ cluster}$$
$$x_i = \text{each point in the } k^{th} \text{ cluster}$$
$$c_k = \text{the centroid of the } k^{th} \text{ cluster.}$$

The sum of the squared errors E_K^2 measures the error representing the points in the total k cluster, which is shown in the Equation 3.2. Therefore, the target of the K-Means algorithm is to calculate the minimal E_K^2 for describing the unchanged assignment and centroids in the computation.

$$E_K^2 = \sum_{k=1}^{K} e_k^2 = \sum_{k=1}^{K} \sum_{i=1}^{n_k} \left\| x_i^{(k)} - c_k \right\|^2 \tag{3.2}$$

where

$$K = \text{number of clusters}$$
$$n_k = \text{number of points in the } k^{th} \text{ cluster}$$
$$x_i = \text{each point in the } k^{th} \text{ cluster}$$
$$c_k = \text{the centroid of the } k^{th} \text{ cluster.}$$

3.3.2 Applications of K-Means Clustering for Fault Diagnosis

Since the first use of K-Means clustering algorithm for multivariate observations by MacQueen [Mac67], nowadays, the algorithm has been applied into various areas, including specific cancer diagnosis in medicine [TT15], vehicle route optimisation [gJmH12] and image processing [NKF$^+$16], etc. In the field of power engineering, the related applications contain condition monitoring of partial discharge in transformers [MSJ04], electricity load modelling [KHML13] and anomaly pattern recognition [OTYB12].

[DWZ+15] presents an method using the K-Means algorithm to classify and recognise the patterns of voltage sags related to incidents in distribution substations. The considered features of large scale voltage sags are duration and amplitude. Concerning the clustering of voltage sags, the results show the probable patterns representing the corresponding characteristics, which belong to possible causes of incidents in the substations. For patterns grouping shorter sag durations and lower sag amplitudes, it may indicate the weak link of the power system with associated substations. The method intensively supports condition assessment as an efficient pre-analysis of voltage sags by clustering the patterns.

Although the K-Means clustering algorithm sometimes can correctly group similar events or objects together, the classification of the patterns lack reasoning. In addition, the number of centroids should be defined initially before the clustering. A wrong choice in the number could cause a misclassification of the real patterns.

3.4 Knowledge-Based (KB) Systems

Knowledge-Based Systems (KBS) are a major part of AI techniques. Generally speaking, a KBS is a computer program which utilises or generates knowledge from data, information, and prior knowledge to solve complex tasks [Lug09]. The system is capable of understanding the information obtained and is able to make decisions based on reasoning with the information or knowledge, whereas the traditional computer systems cannot learn and know the information they capture and process [RN10].

'Knowledge' can be defined as the information *about* information that people use to solve problem. It links a result from observation or analysis, to experiences, laws, and judgments. Therefore, knowledge can consist of facts, models, concepts, heuristics, and examples. Meanwhile, knowledge may also be specific or general, fuzzy or exact, and procedural or declarative. As a result, knowledge can be clas-

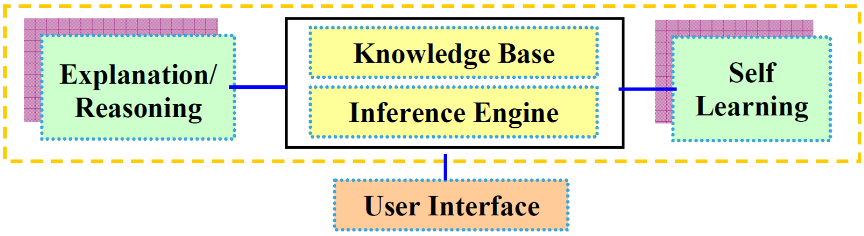


Figure 3.6: Architecture of a KBS [AS10]

sified into various types, including domain knowledge, meta knowledge, commonsense knowledge, heuristic knowledge, explicit knowledge, and tacit knowledge [Smi13, AS10]. For example, domain knowledge is to solve the problem in a particular domain with a combination of theoretical understanding of a problem and experts' experience [Rud10].

Two main components are contained in a KBS: a knowledge base and an Inference Engine (IE). Figure 3.6 shows the architecture of a KBS. The IE is a software program that infers the available knowledge in the knowledge base to solve problems. This may include explanations, reasonings, or simulations to assist the decision-making process. In addition, there should be a user interface to modify the KBS manually or inform the users automatically.

In recent times, KBSs are tightly integrated with developments in the fields of health care, social support, and economic analysis. The KBS applications involve diagnosis, prediction, control, planning, design, maintenance, interpretation, etc. It manifests in different types. In this section, three subcategories of KBSs are introduced: rule-based systems, model-based systems, and case-based systems, alongside a description of how they are applied in fault diagnosis in power systems.

3.4.1 Rule-Based (RB) Systems

3.4.1.1 Overview

Rule-Based (RB) systems (or *production* systems) are one of the simplest forms of AI. They are also the most intuitive and widely-used KB systems in practice and experiments. A RB system uses rules to represent domain knowledge and encodes the rules into systems for problem solving. Instead of representing knowledge in a declarative way describing the truth of things, a set of heuristic rules which are extracted from human knowledge that tells the user what to conclude or what to execute in different situations [Lug09].

Typically, the rules are expressed in the form of if-then (or when-then) statements (called IF-THEN rules or production rules), where“If” (or “When”) is a formula defining the conditions of when the rule can be applied, and “Then” is the effect of fulfilling the particular conditions in the rule, this could be actions or logical decisions. A general rule can have multiple conditions jointed by logical operators (e.g. AND, OR) which precludes a single or multiple consequences [Lig06, GA11].

The structure of a RB system is similar to that of a KB system, which is given in diagram in Figure 3.6. The general architecture consists of following elements:

- Rule base, contains the IF-THEN rules representing domain knowledge. Usually, the rule base is managed independently from the main system to facilitate convenient modifications and extensions to the rules.

- Database, consists of predicate input data for analysis (also known as facts) that match the “IF” parts in the rule base.

- Inference engine, responsible for manipulating the input facts to match with associated rules and carrying the required reasoning to reach a solution (e.g. determine the firing of rules).

- User interface, provides the communication between the system and the users. The design of interface can be in various forms (e.g. menu-driven, graphical-based) through the processing of natural languages.

- Explanation subsystem, analyses the processes of reasoning performance of the system and explains it to the users, giving any conclusions about the facts and rules used by the system.

When the inference engine fires a rule by comparing the facts in the database, the inference chains are obtained. The chains indicate how the system reached its conclusion by applying which rules. There are two main strategies of executing: forward chaining and backward chaining [GA11].

Forward chaining is data-driven, that is to say, given a set of facts in the working memory, relevant rules are applied to generate new facts or conclusions until the expected goal is achieved. This way is always applied in the system where all preliminary facts are prepared [Rud10]. Drools is a common example of forward chaining engine [Red13]. This thesis will discuss the details of Drools application in later chapters.

Conversely, backward chaining is goal-driven, focusing on stating a hypothesis goal and using an inference engine to find the linked proofs. It starts by selecting the rules with conclusions matching the goal in the working memory, then the identified conditions of the rule become sub-goals for matching with other rules' conclusions. The backward work will not be stopped until the facts or provided information can satisfy all the sub-goals [GA11, Lug09]. In the backward chaining systems, Prolog engine is an example for reasoning, details of which can be found in [Bra01].

The choice of forward or backward chaining depends on how the domain experts solve problems. If the system holds all possible information of the problem statement for inferring potential goals, forward chaining is preferable, because it is easy to logically inference the goal with determining the initial conditions However, if the problem data is not easily available and facts must be tested to find a ground for proving the hypothetical solution, then the backward chaining can be applied for the systems.

3.4.1.2 Application of RB Systems for Fault Diagnosis

For power engineering, the development and research of RB system originated in the 1980s, and since then, it was widely used in the design, operation, control, and protection of power systems [M. 94]. The applications cover load forecasting [KEDH02a], protection setting and coordination [EAJMB05, LL96, LYYJ90], service restoration and remedial action [KV91, PL97, TCK+00], and classification and analysis of power system events [SBG02].

Along with the successful application of RB system in many aspects of power system, fault diagnosis was studied to apply RB systems for increasing network reliability. Recent research work of RB system focuses on power system fault diagnosis including the fault identification and location estimation [KNUF92, MIK+95, FK86], condition monitoring of operation and protection equipment [HPF12, RMJ10a], and analysis of general incorrect operations [MRJ97], etc. In this subsection, the applications of RB systems related to fault diagnosis in high voltage equipment (e.g. power transformers) are discussed.

Partial Discharge (PD) is a phenomenon within high voltage insulation in power system equipment. A variety of techniques are used to detect and diagnose PD data. [RMJ10a] presented a KB system offering a generic approach of analysing phase-resolved patterns with identified associated physical PD processes to diagnose faults in gas insulated substations and transformers. Instead of analysing PD data independently using knowledge pertaining to Ultra High Frequency (UHF) data, the research highlights an application of diagnosing defects by recognising both the UHF data and IEC 60270 data, based on the performance of the knowledge captured from UHF phase-resolved patterns. In this developed KB system, the fault diagnosis is divided into five stages to identify a defect location from describing phase-resolved patterns and matching the associated PD behaviour. In order to match related defect characteristics and classify defect, semantic network models were built to represent experts' knowledge and a set of IF-THEN rules are implemented into a RB system to support classification. The rules are invoked and reasoned with forward chaining strategy, executed by the Drool engine. In this work, Figure 3.7 is an example of transforming semantic network model into the IF-THEN rule for detecting the issue of space charge.

Through the direct and useful conversion from a semantic network, RB systems show a powerful solution for assisting engineers to solve problem automatically. However, the RB systems also suffer from some weaknesses, this will be detailed in the later subsection 3.5 discussing the selection of the suitable AI techniques for this thesis work.

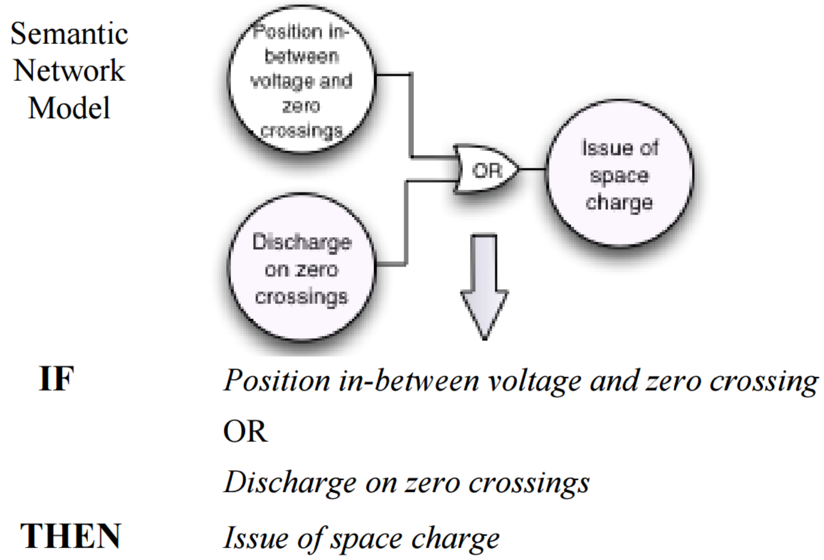


Figure 3.7: Conversion of semantic network model into IF-THEN rule [RMJ10a]

3.4.2 Model-Based (MB) Systems

3.4.2.1 Overview

A Model-Based (MB) system is another well-established application of KB systems for problem-solving. It utilises in-depth knowledge to model the behaviour founded on the specification and functionality of a physical device or system [Lug09]. A created model in the system is based on the theoretical behaviour of a device, which should simulate the ideal device operation or the expected device operation during particular fault conditions. Due to the two types of theoretical behaviour of a device (i.e. the ideal behaviour under normal conditions and expected behaviour during faults), the approaches of Model-Based Reasoning (MBR) are divided into two categories: consistency-based and abductive-based MBR [PW03]. Figure 3.8 illustrates the basic principle of MBR.

For the consistency-based MBR, the model predicts the ideal behaviours (i.e. predictions) when devices or system operates under normal conditions, then the observations (i.e. observed physical devices' behaviours) are compared with the predictions. The identified discrepancies can be used to diagnose fault occurrences in the physical components. The process of matching and identification in the MB systems is supported by a reasoning engine, which is responsible for the control of

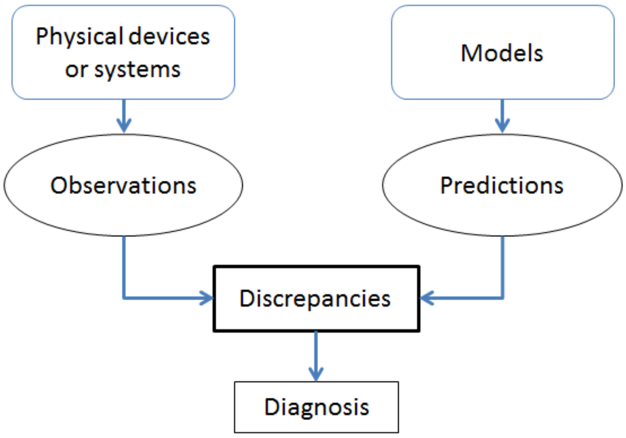


Figure 3.8: The basic principle of MBR [DMM03]

data flow and comparison between the simulated results and measured behaviours [MDDM03]. With respect to the abductive-based reasoning, it focuses on building the model with fault behaviours, the comparison provides the method to identify discrepancies where the observed behaviours matching fault signatures in the models. Generally, the term "MBR" used in typical systems is most referred to the approach of consistency-based reasoning [PW03].

However, no matter what kind of approach, the task of MB system is to tell the experts what to expect, what the differences between the observations and predictions are, as well as how to identify a fault with the discrepancies. The reasoning includes [Lug09]:

- A description of each component of a device, i.e. the simulated behaviour.
- A description of internal structure of the device, which represents the interconnection between each component. The simulated interactions assist the identification of discrepancies and required diagnosis.
- The diagnosis of specific problems by analysing the discrepancies from the model's predictions and observations of actual device's performance. Typically, the diagnosis depends on the measurements of the model's and device's inputs and outputs.

Many MB systems are designed and built with rules to reflect the causality and functionality of a device. The reasoner can find the most probable fault matching the observed system behaviour to ensure the success of MB system operation. Meanwhile, additional rules can be used to describe the modes and their interconnections. But, unlike RB systems, if a problem instance does not match the heuristic rules in their rule-bases, the problem will not be correctly diagnosed even though an in-depth analysis would find a solution to identify the instance. Models in the MB systems allows for a wide range of reactions to the input stimuli, containing some that are measured but not considered during the implementation. Therefore, this is also one advantage of MB systems; it improves the reliability of attempting to address the situations that RB systems encountered. Due to this benefit, MB systems are particularly applied to systems involving complex interactions and large numbers of components.

3.4.2.2 Application of MB Systems for Fault Diagnosis

The earliest MBR applications appeared in the 1970s, and now MBR has been widely used in power system for fault diagnosis, especially in the domain of power system protection. Models of electric equipment (e.g. transformers, protection relays, etc.) are available for the automatic analysis of condition monitoring [AHF+98], protection operation validation [MDM+96, BMM+98], fault location and diagnosis [LFST94, BDF+93], alarm processing [EDM+13], etc.

[AHF+98] presented an application utilising the MBR approach to assess the condition monitored in power transformers to identify faults. This approach used adaptive thresholds which can be altered based on the working conditions of transformers. The monitored data includes temperature, load, over-voltages, gas and moisture in oil. Then the MBR engine compared past output values of various built-in sensors with the predicted results from mathematical models. Since the transformer fault is sometimes caused by a number of reasons and the characteristics are not simply linear, the developed MB application could only support a signal indicating the existence of fault conditions in the transformers without identifications of fault type and location. In order to solve the problem, two possible solutions were discussed. One is to use RB systems for fault diagnosis, well-defined rules could assist the experts to directly target corresponding faults based on provided measured conditions; the other is a case-based approached. The latter technique will be reviewed in subsection 3.4.3. However, both of the methods were limited due to the lack of available data of normal and fault conditions and the lack of experts' knowledge of transformer faults.

The Decision Support System (DSS) proposed within [MDM$^+$96] is integrated with the MBR approach for protection operation validation. It concentrates on automating the experts' work of analysing protection operation after a fault occurrence. This system contains two models for SCADA alarm processing and fault diagnosis by utilising a General Diagnostic Engine (GDE) and SCADA data. The GDE is developed for the assessment of protection operations by modelling the entire protection schemes of correct behaviours [KW87]. In the DSS, the consistency-based reasoning method is applied, the expected behaviours from current transformers to circuit breakers are validated against observed behaviours, i.e. circuit breaker open and current values. The use of current measurement is not only for signalling the response of the protection device, but also for verifying the behaviour reaction to the faults. Although this system could provide the functionalities of validation for protection operations, it always cannot offer the information of specific failure components, for example where some faults occurred in communication links or in basic elements which are not modelled. Therefore, model-based applications have their own limitations for providing accurate performance under some circumstances.

3.4.3 Case-Based (CB) Systems

3.4.3.1 Overview

Apart from heuristic rules and theoretical models to solve problems with experts' knowledge, another powerful strategy is Case-Based Reasoning (CBR) which uses human knowledge to address new situations by reasoning using explicit cases of past problems or solutions [Lug09]. When Case-Based (CB) systems run into a problem faced before, the reasoning will seek and perform the matched solution against previous experienced case in the case base. Meanwhile, when the CB systems encounter a similar scenario, the CBR will find similar cases refer to their experience and a solution can be developed based on previous provided solutions without reasoning from first principles. Therefore, the solutions may reflect the results of previous search-based successes or failures, or may be processed through experts' knowledge engineering.

One important aspect of CBR is remembering; once the new problem is solved by the searched or developed solution, the particular case will be saved in the case base for solving new problems in the future. These added cases will improve the reliability of the CB systems by giving more accurate results through the CBR [Xu94]. The common processing structure of CBR contains four main steps, which are illustrated in Figure 3.9 [Lug09]:

1. *Retrieve*: after assessing the current case, the appropriate cases and solutions are retrieved from the case base. This performance is based on the similarity, which is determined by their common features. For example, if two patients share a number of common features in their medical drugs and histories, then they probably have the same disease. So, typically, the cases are indexed by their significant features to increase the efficiency of retrieving.

2. *Reuse*: if the retrieved cases completely matches with current problem, then the solution can be directly reused. Usually, the retrieved solution is required with modification to adapt better to the new problem. The reasoners need to transform the saved solutions into suitable operations for current problem. For example, even if the two patients have the similar medical drugs and histories, some special circumstances can make the patients have different diseases, so the doctors should make "adjustments" for the two patients in his/her mind.

3. *Revise*: sometimes, a modified solution may not lead to a satisfactory result. In this case, the solution should be revised before it is applied to the problem

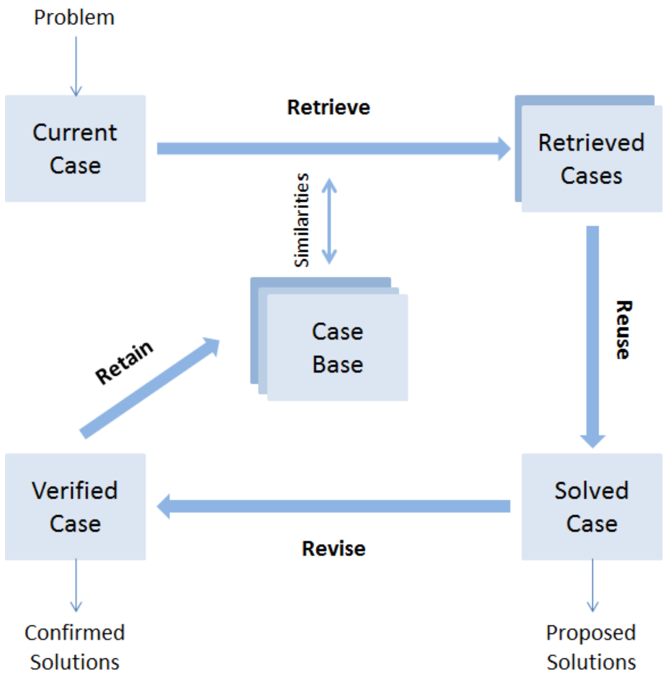


Figure 3.9: The processes of CBR [AP94]

or after the solution has been applied. That is to say, previous steps will be iterated, i.e. more cases retrieved and the solution adapted.

4. *Retain*: if the solution was verified as correct, then the solution will be saved with a record of success or failure. The new case will be updated into the case base for future use.

From the description of CBR processing, the main advantages of applying CB systems can be easily realised [Xu94, Lug09]:

- The system can be designed and developed by encoding cases for solving problems in the specific domain. It is not necessary for understanding the knowledge itself. Only the cases coded and the performance results are keys for the CB systems to focus on. This simplifies the knowledge engineering process and can be particularly suitable for situations where domain knowledge is lacking.

- Compared to the RB or MB systems, CBR does not need complex reasoning for finding or developing a solution to solve problems. When an appropriate case is searched in the case base and totally matched with the current situation, then the previous solution can be used directly to address the problem, which is more effective than other strategies.

Nevertheless, the CBR technique has its own shortcomings. This is because the reasoning is not required when there is abundant knowledge, the CB systems cannot always provide sufficient explanations of reasoning for the results when a solution is found. Furthermore, with the increase in the number of cases in the database, the computation of retrieving the entire case base for a solution becomes more time-consuming. Thus, in order to determine the similarities among the cases, an efficient indexing and similarity matching algorithm is necessary to be developed in the CB systems.

3.4.3.2 Application of CBR for Fault Diagnosis

Although CBR is a relatively new AI technique, there are numbers of successful applications in the academic and commercial domain. Most of them are used in people's infrastructure services, such as medical diagnostics and intelligent voice services. For the area of power systems, CBR is less widely applied. [WSM+01] proposed Design Engineering Knowledge Application System (DEKAS) to assist the design of protection systems in transmission network by developing new cases with utilising experts' knowledge and similar past designs. [QGY06] discussed the application of CBR to classify the inception fault of the transformers. The fault diagnosis focuses on retrieving the most similar case matching the symptoms of the target case from the case base, and then defining the fault classification.

At present, MB condition monitoring and fault diagnosis can improve system's reliability and availability, and reduce the maintenance costs by enabling the detection and identifications of incipient and repeat faults very reliably. However, for some equipment such as circuit breakers, the MB systems cannot offer a cost-effective application for diagnosis feasibly. For example, [SMF01] presented a methodology of combining the MB and CB approaches to diagnose the faults within circuit breakers, which could utilise a low-cost computer but provide high-quality performance. The technique is mainly divided into two parts: preparation and diagnosis. The MB and CB diagnosis are responsible for each task respectively, the model simulates the fault modes of circuit breaker off-line and the search module takes the evaluated results (i.e. cases) to match the most likely diagnosis on-line. The contrast of similarity between diagnostic cases depends on the features defined and derived from the monitored quantities in the system (e.g. contacting time, position and velocity). Through the use of a combination of MBR and CBR, the system could produce a diagnosis for the current condition of circuit breakers under a low-cost computation. However, future work is still required to improve the performance, for example, the monitored parameters of specific device should self-adapting; advanced methods for pre-processing of sensor data in order to obtain more significant additional features are necessary.

3.5 Selection of AI Techniques for Automatic Fault Diagnosis and Prognosis

In the previous sections, this thesis overviewed various AI techniques and their applications in the field of fault diagnosis in the power system. The purpose of evaluation is to search the most suitable approach of fault diagnosis and prognosis with using the analysis of SCADA alarm data and overhead line network data, which is the core task in this research work. A KB system is considered for automatic data analysis by invoking RB generated from experts' knowledge, data mining, and pattern recognition clustering algorithm. In this section, the reasons for selecting such an approach as opposed to alternatives are discussed.

3.5.1 Selected Approach: RB Strategy and K-Means Clustering

In this research work, as introduced in Chapter 1, the automatic fault diagnosis with analysis of PMAR device faults and abnormal network fault events is supported by the use of experts' knowledge. Consequently, a KB approach is the most suitable and convenient AI technique that can be adopted to address the targeted task of this research work. Table 3.1 summarises the strengths and weaknesses of the RB, MB and CB approaches with consideration of the specialities of this research work.

	Strengths	**Weaknesses**
RB	• Precisely and naturally represent experts' knowledge of experienced faults	• Lack of human common sense on some decision makings
	• Flexibly access to update the diagnostic and prognostic rule-base	• May fail if a fault scenario (e.g. particular PMAR device fault) is not encoded in the rule-base
	• Excellent explanation facilities	
	• High level of automation to simply and easily verify and validate	
MB	• Incorporation of numbers of functional models of physical systems or devices, identifying the fault events accurately	• Will not work if the component or link is not modelled
	• Capability of predicting fault events through simulation of abnormal activities	• Different PMARs should have different model settings
CB	• Suitability for incomplete knowledge domain	• Slow at handling large case bases
	• Convenience of diagnosing the specific PMAR device faults	• Not suitable for identifying circuit faults due to the various conditions
	• Incremental learning	
	• Ease of knowledge elicitation	

Table 3.1: Summary of the strengths and weaknesses of the RB, MB and CB approaches for automatic fault diagnosis

The fault diagnosis of PMAR devices and network conditions is conducted based primarily on the distribution network operator's (i.e. SPEN's) engineers' knowledge. In practice, the faulted conditions of PMAR devices and distribution networks are specified and defined by the manufacturers and the network operator, which are mostly written in the form of rules. So moving the ready-made rules to a RB system is a natural and suitable approach. Meanwhile, the RB approach can offer explanations regarding the output reasoning when the suggested solution matches with detected conditions. This is a precise and convenient way to support the engineers with identifying the causes and taking positive actions. Furthermore, the separate reasoning process and data manipulation will not affect the update of rule-base in the whole system, engineers can flexibly access the rule-base and manage the rules with requirements.

For the MB strategy, although it can incorporate different fault models to detect and identify the specific fault scenarios within the PMAR devices, the links between components is not always easily realised and modelled. Furthermore, the RB system can diagnose the fault scenarios more directly through clear rules than MB systems, which finds the discrepancies between the simulated results and observed behaviours. The main inconvenience of applying MB systems is with the difficulty of modelling the settings of PMAR devices, some specific PMAR devices could require particular configurations to keep the correct functionalities; this could lead to larger volume of work. For the CB strategy, although it is possible to define different cases to specify the settings to diagnose PMAR device faults, it needs significant effort to classify the fault conditions on the overhead line, which require more details to define various fault types. Such an approach will generate a range of similar cases, imposing a burden on the system's computation.

There is a common problem existing in these three strategies, that is, the system may be fail if a fault scenario is not defined or modelled. For this situation, the RB approach is easier and quicker to add or delete rules when compared to the other two strategies that must update their knowledge-base. Hence, the RB approach is adopted in the implementation of DSS for automatic fault diagnosis of PMAR device and abnormal network conditions.

With respect to support control and protection of distribution networks, the automatic fault prognosis will predict future potential fault events by analysing current network conditions. The predictors (i.e. the fault features which are capable of forecasting future fault activities) and predictions (i.e. the attributes of potential fault events, such as time of occurrence, locations, etc.) are uncertain. Unsupervised K-Means clustering algorithms can be utilised to efficiently group the fault activities with similar predictors and predictions. Then, the thresholds of some clusters (which contains the features that can be used for particular fault diagnosis and prognosis) could be generated into rules, and implemented in the KB system to indicate future network faults. Also, the K-Means clustering results are easy to interpret for classifying and recognising patterns.

Despite ANNs being used for classification and forecasting applications, this machine learning technique generally involves training a set of data with fixed outcomes when the same network data are inserted. This data is not readily available in the case of this research work. So this can be difficult to provide suitable predictive rules without adding desired values in the technique. As a result, this thesis selects K-Means clustering algorithm and data mining to offer satisfactory solutions for automatic fault diagnosis.

3.6 Conclusion

AI is a broad discipline with great development potential, which intends to solve problems leveraging the efficiency of machines whilst retaining the intelligence of human beings. This chapter reviewed the AI techniques that are widely used in the power systems for fault diagnosis and condition monitoring including: RB systems, MB systems, CB systems, K-Means clustering, and ANNs machine learning techniques. Based on the comparisons between the advantages and disadvantages of these techniques, a KB system has been chosen for the automatic fault diagnosis and prognosis by invoking a rule-based approach, which is implemented using experts' knowledge, data mining, and a clustering algorithm. Although the other AI techniques have the potential to solve the task of the research work, they may require extra effort to attain the same level of suitability and efficiency, such as the MB approach should build different models in response to various PMAR settings, or the CB approach could revise much cases and solutions to identify underlying circuit faults. The following chapters discuss the design and deployment of DSS with implementation of the KB system, which also verifies the appropriateness of the selected approach.

Chapter 4

The Design and Development of a Decision Support System for Automatic Fault Diagnosis and Prognosis for Distribution Automation

4.1 Introduction

Chapter 1 introduced the core novelties of the research work, which are:

- The identification of the root causes of the anomalous fault problems (i.e. identifying the degradation of devices and circuits) with automatic analysis of SCADA alarm data and PMAR data.
- The prediction of the evolving faults (i.e. providing 'early-warning' of PMAR operations) to help mitigate customer service interruptions and improve system reliability with an in-depth diagnosis of PMAR data.
- The employment of a KBS for automatic analysis of fault events, and anomalous activities by applying appropriate AI techniques.

These DSS capabilities were developed by applying the selected approaches discussed in Chapter 3. Therefore, details of the proposed methodology and the operation of the DSS are presented.

This chapter introduces the manual process of daily protection operation analysis by SPEN's engineers, along with a description of the involved data types. The chapter then continues to detail the design and operation of the entire automated DSS processes for fault diagnosis and prognosis, which includes an integrated visualisation tool. This tool supports the end user with advanced fault diagnosis and prognosis functionalities to enable the main objective of supporting the end user's decision-making process.

4.2 The Manual Process with SCADA Alarm Data

4.2.1 OHLs' Unsolicited Openings

Overhead Lines (OHLs) usually deliver electricity in remote areas and distributed areas of low population density. Difficult terrain and longer distances between DNO depots and interrupted customers sometimes are the main reason of delayed supply restoration and network reconfiguration to these areas. [SP 12a]. When a wide-area blackout occurs, these remote areas have a lower priority of restoration compared to areas supplying higher population densities. Meanwhile, over 50% of OHLs' faults are transient, which occur due to factors such as storms, animals or lightning. Although these short-lived faults appear and almost immediately clear themselves, the effect of a transient fault on OHLs can cause protection equipment to operate. This can increase the CMLs and restoration costs. Therefore, OHLs require a more reliable protection system to stabilise the supply within regulatory requirements. Consequently, SPEN has implemented microprocessor-controlled OHL PMAR technology to overcome these issues [SP 12a].

As introduced in Chapter 2, when a fault is transient in nature, 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; when a permanent fault occurs, the reclosure attempt will not succeed in isolating the fault, and the PMAR will 'lockout' on its third attempt. Following a 'lockout' 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 cleared from the circuit or isolated from the PMAR in question.

With respect to the class of semi-permanent faults, the degradation of OHLs causes frequent short-term supply interruptions, which leads to trips or regular lockouts of the PMAR, called 'nuisance tripping'. Not only do these semi-permanent faults affect the quality of daily electricity service, such as 'nuisance tripping' due to frequent transient fault activity, but they may also evolve into more serious permanent faults that result in significant outages. Such outages incur more expensive repair costs, as well as more CMLs and CIs.

If the PMAR fails to remain closed after a manual remote reclosure following a lockout, it is not conclusive evidence of a permanent fault. In some instances, where fault conditions remain after the PMAR locks out, the control engineer will continue with reclosure attempts via telecontrol for a period until the fault condition is manually cleared or it then becomes apparent that a permanent fault most likely exists on the line. However, there remains some uncertainty surrounding the cause of PMAR operations which do not culminate in the definitive identification of a permanent fault, i.e. operations that either do not lead to a lockout or where the control engineer ultimately recloses via telecontrol after the PMAR has experienced a number of lockouts in relatively close succession. In such cases, there is rarely any understanding of the root cause of the initial and subsequent PMAR trips. This undiagnosed PMAR activity is referred to as an 'unsolicited opening' (UO).

4.2.2 The Manual Process of Analysing Semi-Permanent Fault Activity

From a technical and/or asset management perspective, the frequency of UOs on a particular circuit could provide some indication of deterioration in a circuit's (or indeed an individual PMAR's) performance and underlying condition. For example, the cracked insulator previously mentioned, 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. Thus supply to the circuit is restored for a time, until the scenario repeats itself at some point 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, in fact, may not be the case. Also, from a customer service perspective, the 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, and also exposing the DNO to regulatory penalties.

At present, for the purpose of identifying the particular areas or points on a circuit affected by frequent UOs to support engineers' further analysis and decision makings (i.e. diagnosis of causes of UOs and suggestion of appropriate actions), SPEN engineers are required to periodically (and manually) review the number of UOs that occur on the network, in an attempt to ascertain the degree to which the network is affected by this activity. This requires the careful consideration, alignment, and analysis of a number of separate data sets, which are stored in different repositories, in order to build a coherent picture of circuit behaviour and the potential problem. The utilised data sets are:

- Morning report: is the daily report generated manually by the control room team, including the relevant UO operation activity (e.g. specific PMARs name, component code, etc.) captured from SCADA system information.

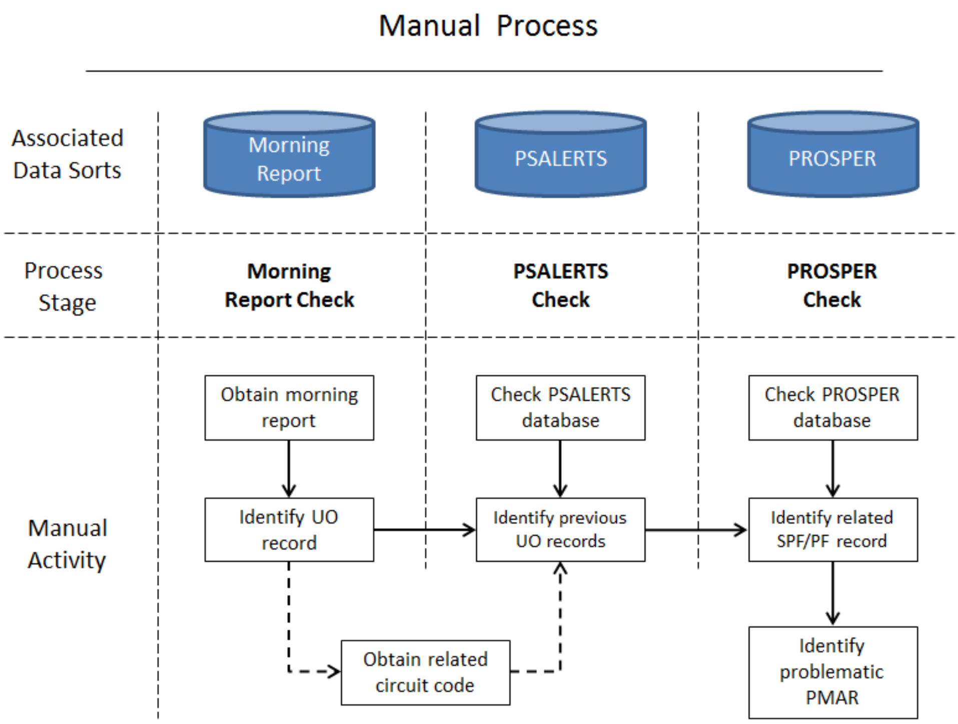


Figure 4.1: The manual process of analysing UO (semi-permanent fault) activity

- PSALERTS: is a data set that archives SCADA alarms which include tripping information (e.g. test trips, UOs, etc.) associated with the particular circuit details (e.g. activity log time, specific PMARs name, circuit code, etc.).

- PROSPER: is a database that includes causal information related to repaired faults, which are generated manually by the maintenance staff. This information contains the causes of faults, the fault clearing time, relative affected equipment, such as the corresponding PMARs name.

Figure 4.1 shows the manual process currently undertaken for identifying problematic PMAR by analysing associated data types. The process contains three stages: morning report check, PSALERTS check, and, PROSPER check.

As illustrated in the diagram, the engineer's focus is identifying the problematic PMAR through checking the frequent UOs of the relevant circuits. First, the corresponding circuit code related to the identified UO activity is retrieved. Then, analysis of previous frequent UOs of the same circuit (by checking records in PSALERTS) is conducted. This is to identify events that may proceed a resultant fault which was repaired manually and recorded in PROSPER. Finally, engineers diagnose the causes of faults related to the UOs. Through the manual process, the analysis may provide a picture of circuit behaviour and underlying conditions, and with the identification of problematic PMARs, maintenance staff can undertake repairs preceding the potential outages raised from these PMARs.

However, due to the limited information provided in PROSPER, the UO analysis performed (manually) by engineers can only identify problematic circuits and PMAR devices without necessarily focusing on diagnosing the root causes of PMAR operations. In order to obtain a clear view of OHLs' conditions and provide control engineers more accurate decision supports with diagnosing root cause problems and predicting evolving faults, this thesis describes the design and implementation of a DSS that the engineers use, following their manual analysis. The DSS fully automates the existing approach to analysing and quantifying UO activity associated with PMARs and circuits, while integrating fault diagnostics. Furthermore, the DSS also has a predictive capability which would alert engineers to incipient fault conditions.

This diagnostic and predictive capability is presented as a Knowledge-Based System (KBS), with the development of a rule base which mainly makes use of the data captured in the PMAR log files, containing the details of circuit behaviour and PMAR's condition. These data recorded as 'pick-up' activities, including anomaly activities (which do not lead to a trip operation) and fault events (which lead to a trip or lockout operation of the specific PMAR), i.e. the current amplitude of pick-up activity, the affected phases, and the time stamp of occurrence and clearance of pick-up activity, etc., previously introduced in subsection 2.4.3. The design of DSS and implementation of the deployed KBS will be detailed in the following section.

4.3 The Design and Development of Knowledge-Based Decision Support System

This section describes the designed functionalities of the DSS which automates the manual process and extends it with the novel analysis of PMAR log files to support engineers' decision-making with diagnostic and prognostic results. The detailed architecture and activity diagrams of the DSS facilitate the development of the deployed KBS to deliver these proposed functionalities. Additionally, this section introduces how the designed DSS applies the knowledge based techniques to analyse the SCADA alarm data and PMAR log data automatically, and which RB strategy and inference engine are chosen and applied in the implementation of KBS.

4.3.1 The Design of DSS

4.3.1.1 The Designed Functionalities of DSS

With respect to limitations and weaknesses of the manual process introduced in earlier sections, the designed DSS not only fully automates the periodical processing of SCADA alarm data to identify problematic PMARs, but also provides decision support with diagnostic and prognostic analysis of the PMAR log files' data. Figure 4.2 illustrates the proposed functionalities of the designed DSS for supporting engineers' decision-making, which provides emergency suggestions (to be agreed and executed by control engineers) and recommended solutions (combined with consideration of experts' experience). Both decision support interventions deliver the benefits for automatic distribution network protection and operation.

By obtaining and analysing the associated data at sequential stages, the designed DSS will be able to provide:

- the combined information on circuits which are affected by frequent UOs.

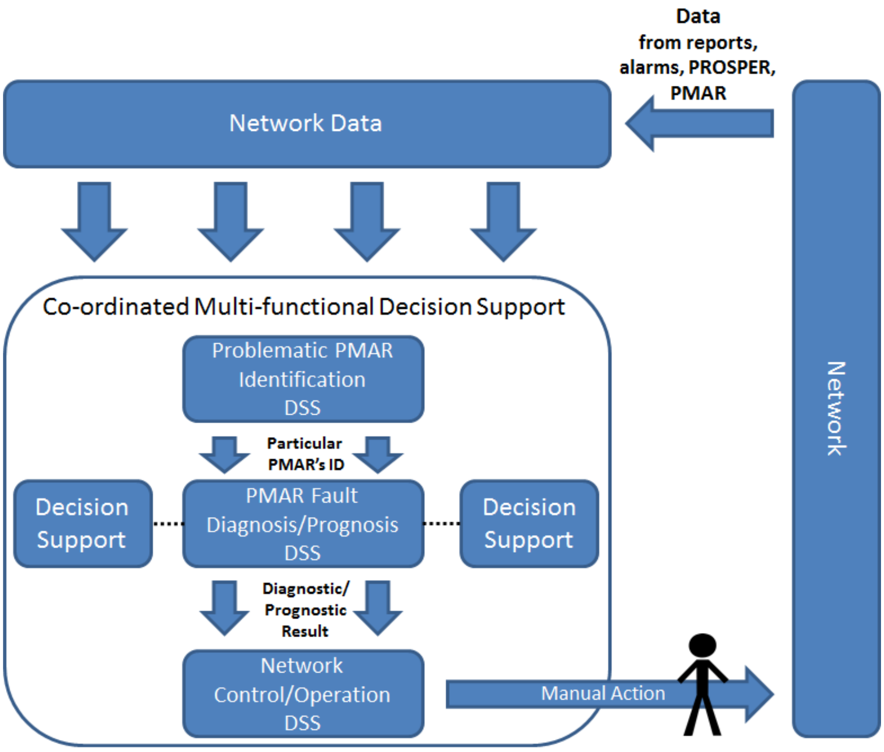


Figure 4.2: The functionalities of designed DSS

In addition, the information indicates the associated PMARs related to the identified UOs, this allows the DSS to focus on the in-depth analysis of PMAR data.

- the diagnostic results of faults related to the PMAR device or circuit's condition. This information may indicate the cause of the failure of PMAR operations due to problems of PMAR devices. It could also identify the semi-permanent faults which lead to frequent UOs affecting the quality of electricity supply, which may be caused by circuit degradation. For example, the detection of permanent damage of PMAR devices or circuits could suggest network reconfigurations to mitigate affected areas or prevent potential outages through DA.

- the prognostic result of the potential PMAR operation in future. This information could offer an early warning to the control room, allowing engineers to take appropriate actions (e.g. dispatch maintenance staff to repair/troubleshoot the weak link in the protection zone of the particular

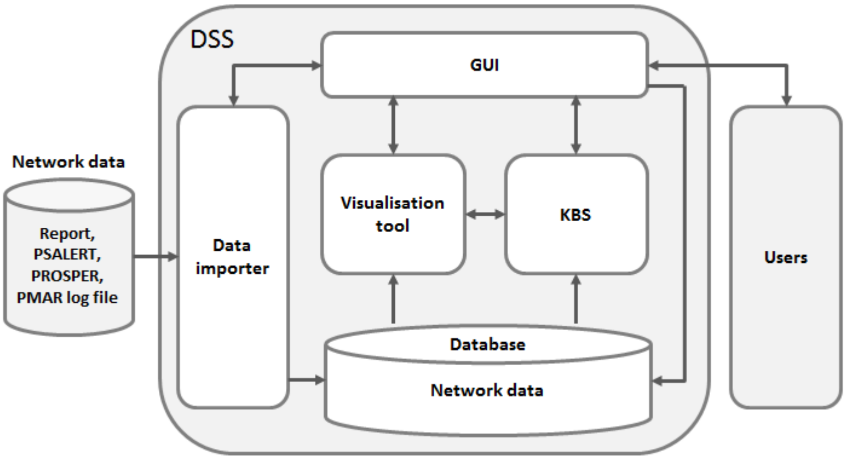


Figure 4.3: The overall architecture of DSS

PMAR) to prevent potential outages, so that can minimise the CML/CIs.

4.3.1.2 The Overall Architecture of DSS

In order to achieve the proposed functionalities, the DSS deploys and develops a KBS to analyse the data for fault diagnosis and prognosis automatically. Also, the visualisation tool was developed as part of the designed DSS to assist in the design of the KBS, along with supporting automated data analysis. The whole DSS system has been implemented on the Java platform [Ora14], where the selected rule engine within the KBS will be described in the next subsection. Figure 4.3 shows the overall architecture of the DSS, including the following main components:

- Data importer: The key element for data interaction between the external network data and internal database of the system. The importer can store the new generated data into the database, and update historical data, all the commands are operated through the developed graphical user interface.
- Database: The inner MySQL data container [Ora17] for holding the imported network data. The data will be pre-processed and classified into various categories for further data analysis.
- Visualisation tool: it allows the users to clearly view the graphical infor-

mation of the circuit's conditions (i.e. fault summary over a period, etc.), this tool also facilitates the tasks of identifying semi-permanent faults for automatic fault diagnosis. Meanwhile, the visualisation tool is used to assist the development of the rule-base for implementing the KBS. Examples of the utilisation of the visualisation tool for automated data analysis will be presented in the case studies of fault diagnosis and prognosis in Chapter 5 and 6.

- KBS: Responsible for performing the automatic fault diagnosis and prognosis by checking the detected conditions against the diagnostic and prognostic rules generated from experts' knowledge, data mining and clustering techniques. By capturing and processing the useful information from the database, the KBS sends the automated analysis results to the visualisation tool for further identification of semi-permanent faults or directly provides the diagnostic and prognostic results to the engineers via the user interface.

- Graphical User Interface (GUI): Allows the information exchange between the user and the DSS. As illustrated in Figure 4.3, the users control all other elements of the DSS through the functional block of GUI. For example, the users can ask the GUI to present some particular PMAR's information through utilising the visualisation tool by calling the associated data from the internal database.

With respect to Figure 4.3, the performance of the DSS is based on the successful implementation of the visualisation tool and the KBS. When the PMAR log is imported or updated in the database, the KBS will automatically invoke the rule-base to diagnose PMAR device faults and detect semi-permanent faults which are responsible for UOs. Simultaneously, the KBS will provide an 'early warning' of a future PMAR operation in order to assist the engineers to reduce the impact on customers before it escalates to more serious and damaging supply outages. All the analysed results will be prepared for posting to the GUI to demonstrate based on users' requests (e.g. view of summary of historical faults of a particular PMAR or circuit), but some results will be given a further analysis, such as trend analysis of semi-permanent faults to support engineers' decision-making.

4.3.1.3 The Activity Diagram of DSS

Figure 4.4 displays the activity diagram of the DSS which describes the functional details of the designed KBS and visualisation tools. The automatic analysis for fault diagnosis and prognosis is based on the identification of problematic PMARs, which continues the work from the automated manual process.

As illustrated in the diagram, the visualisation stage focuses on analysing the details of PMAR's log data (examples shown in Appendix A) after they have been imported into the system. It will automatically infer useful information so as to allow the users to visualise it. This information includes the trips summary, pick-up distributions, and details with selected time periods independently, which offers an insight into the underlying condition of the circuit. Any specific details (i.e. fault type, fault duration, the amplitudes of fault currents and voltage, etc.) of pick-up activities can also be searched for in the visualisation tool. This facility offers engineers the opportunity to find trends in the underlying circuit pick-up activity leading to a PMAR operation, and assists engineers' decision making for fault diagnosis [WMS+17].

The stage of PMAR log check analysis is conducted in parallel with the visualisation stage in the automatic analysis. It contains the functionalities of the diagnosis of PMAR device fault, the detection of the semi-permanent fault diagnosis, and the early warning or prediction of degrading PMAR or circuit conditions. The DSS will automatically identify PMAR device faults based on the detection of specific warnings or messages related to different categories of device faults, which are stored in the PMAR log file. Meanwhile, the extracted pick-up information will be evaluated to determine the existence of the SPF, through identifying the patterns and trends. Finally, the prognostic function in this analysis stage will predict future PMAR operations to mitigate the customer interruptions (which may be raised from SPFs) with matching the appropriate predictive rules [WMS+17].

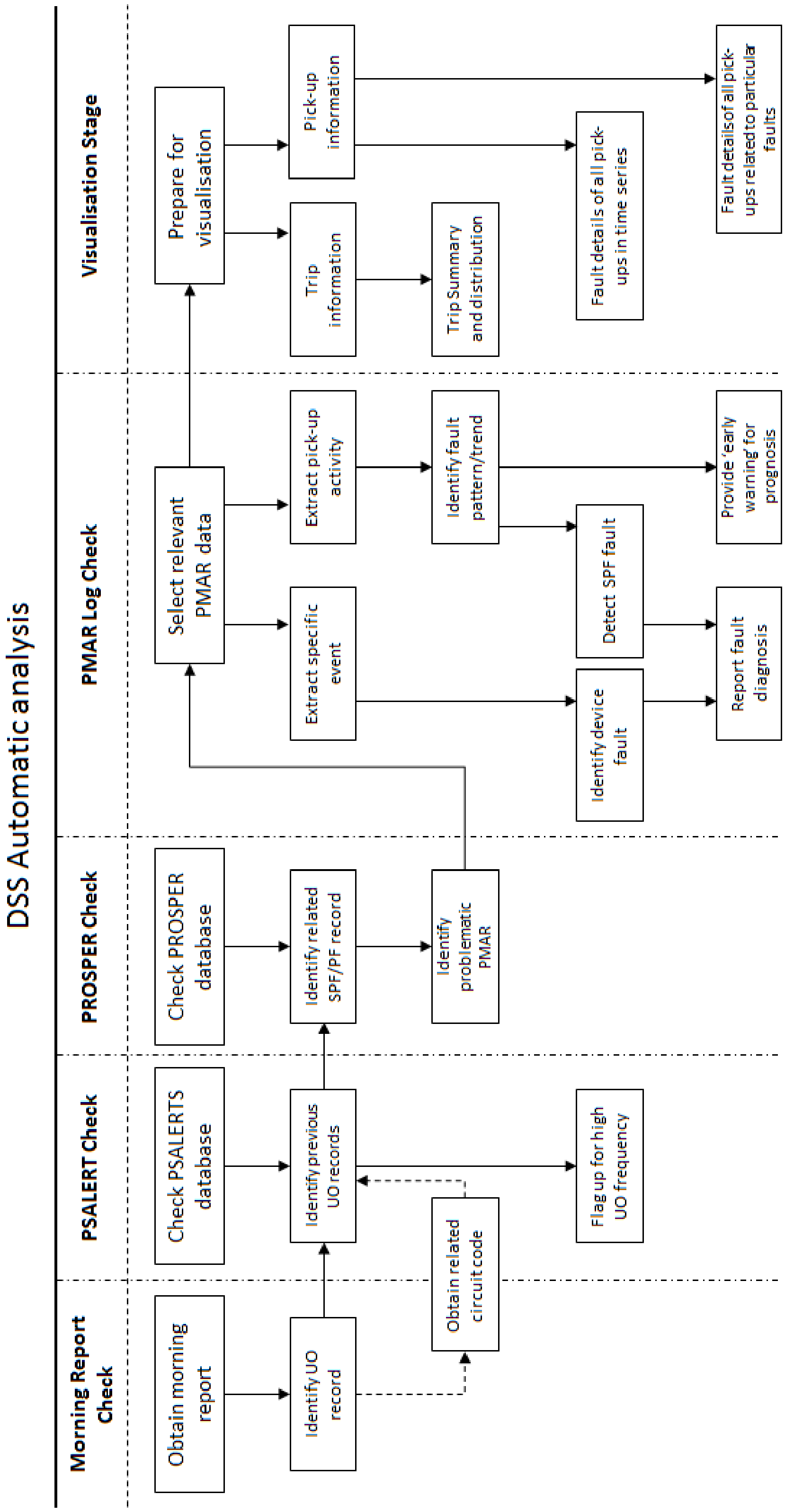


Figure 4.4: The activity diagram of DSS

As for the automated manual process, i.e. the first three stages of DSS automatic analysis shown in Figure 4.4, the designed functions perform what was previously the engineers' responsibility of identifying problematic PMAR devices. In addition, the automated process can also provide alarms when detecting high UO frequency of a particular circuit. The details will be described in the subsection 4.3.2.

At present, the stages of PMAR log check and visualisation are implemented as part of the designed fully automatic DSS, which achieve the functionalities of automatic fault diagnosis and prognosis. For the part of automated manual process, the thesis is more focusing on its design, the extended work of the development of the entire automatic system will be detailed in Chapter 7.

4.3.2 The Automated Process of Identifying Problematic PMARs

With respect to description of the activity diagram (as shown in Figure 4.4) of the designed DSS, this subsection introduces how the DSS automates the manual process of engineers' daily data analysis and prepares for automatic fault diagnosis and prognosis. As indicated, the start of the automated process is the input information retrieved from the daily report, and it is also the start of fully automatic DSS analysis as well. The morning report contains the detected UO activities with their sources (i.e. the particular PMAR device). However, some recorded UO activities in the morning report do not directly show the names of the affected PMAR devices, they provide component aliases which register the UO activity on the circuit. Therefore, in order to obtain the PMAR's name, the automated process will check another data source, named as 'NOJA list', where the PMAR name can be found through searching the component aliases. 'NOJA' is one of the brand names of PMAR devices, which applied by SPEN and deployed in their distribution network.

The details of the automated process to identify the problematic PMARs affected with frequent UOs from checking morning reports, which are expanded

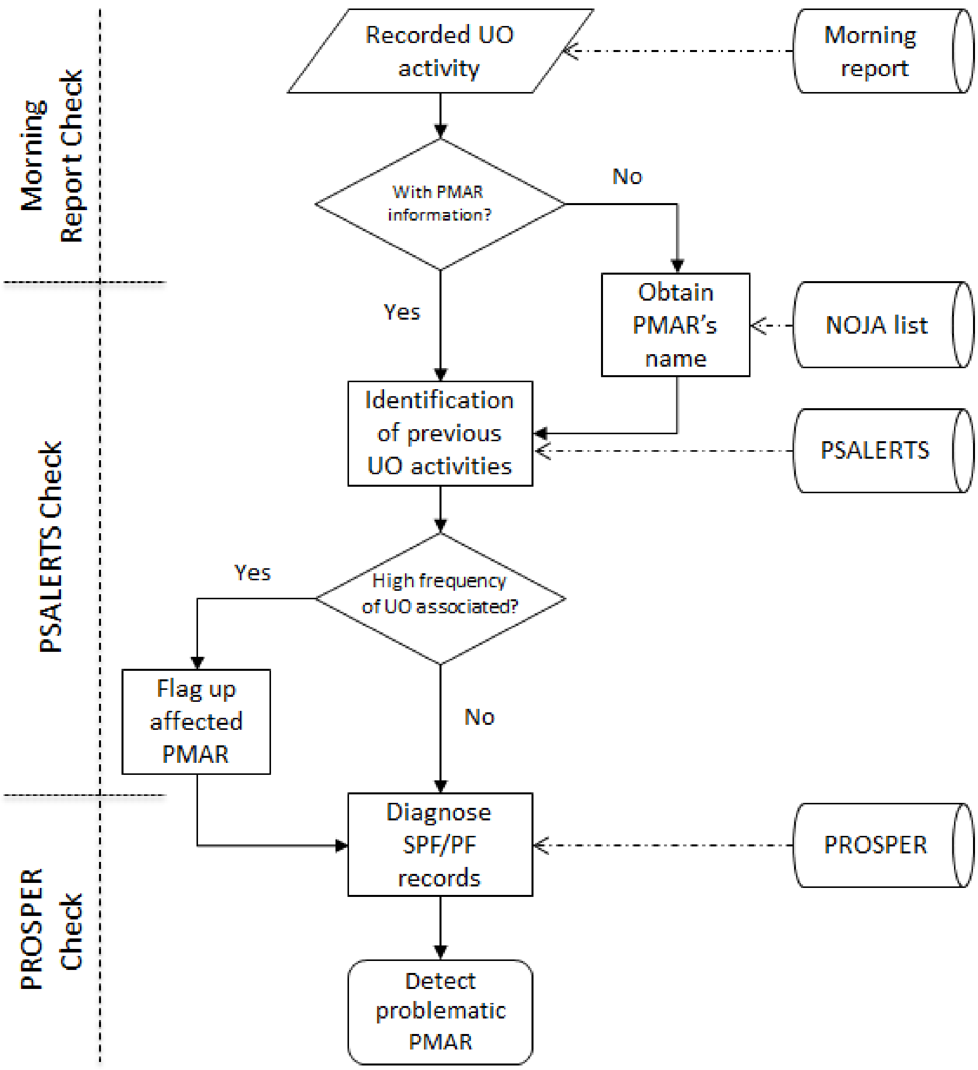


Figure 4.5: The process of automated analysis

in the flow chart in Figure 4.5, and summarised by the following main steps:

1. The morning report provides the recorded UO activity to the DSS which is imported through the data importer, and then the DSS checks the PMAR information with the imported UO activity.

2. If the UO activity contains PMAR information (i.e., the name of the affected PMAR), the DSS will take the PMAR's name and circuit code (i.e. circuit's ID) to search the previous UO activities associated with the particular PMAR. If not, the DSS will bring the component alias to obtain the PMAR's name and circuit code from checking the 'NOJA list' database. Then, the process will drop into the stage of PSALERTS check.

3. When the historical UO activities are retrieved from PSALERTS database according to the provided PMAR information, the DSS will automatically calculate the frequency of UO activities and check the threshold to decide whether to raise an alarm of the seriously affected PMAR with identification of high frequency of UOs. The threshold is preset by the control engineers using their analytic requirement.

4. If the system does not detect a high frequency of UOs with the related PMAR, the process will directly proceed the stage of PROSPER check; if the high frequency of UOs exists, a warning message will be generated and provided to the users through the GUI before the PROSPER check.

5. With the provided circuit code and time stamp of the recorded UO activity, the system searches the manually recorded fault descriptions corresponding to the UOs. If the fault potentially resulted from asset deterioration (e.g. cracked insulator, broken insulation, or broken pole, etc.), the system considers the circuit or the relevant PMARs as affected by potential semi-permanent faults, and reports the problematic PMARs to the engineers and extracts the related data for further automatic analysis.

Due to the PROSPER database it always cannot provide the useful information of causes of UOs registered by the related PMARs. As shown in Figure 4.4, after the process of PROSPER check, the DSS will automatically continue the in-depth analysis of associated PMAR data for fault diagnosis and prognosis within the developed KBS. The following subsection discusses the design and development of KBS which focuses on analysing PMAR data for automatic fault diagnosis and prognosis.

4.3.3 The Development of RB KBS

As discussed in Chapter 3 regarding the selection of the appropriate AI techniques for fault diagnosis and prognosis, the fault diagnosis of PMAR devices and evaluation of circuits condition is primarily based on the specifications of manufacturers

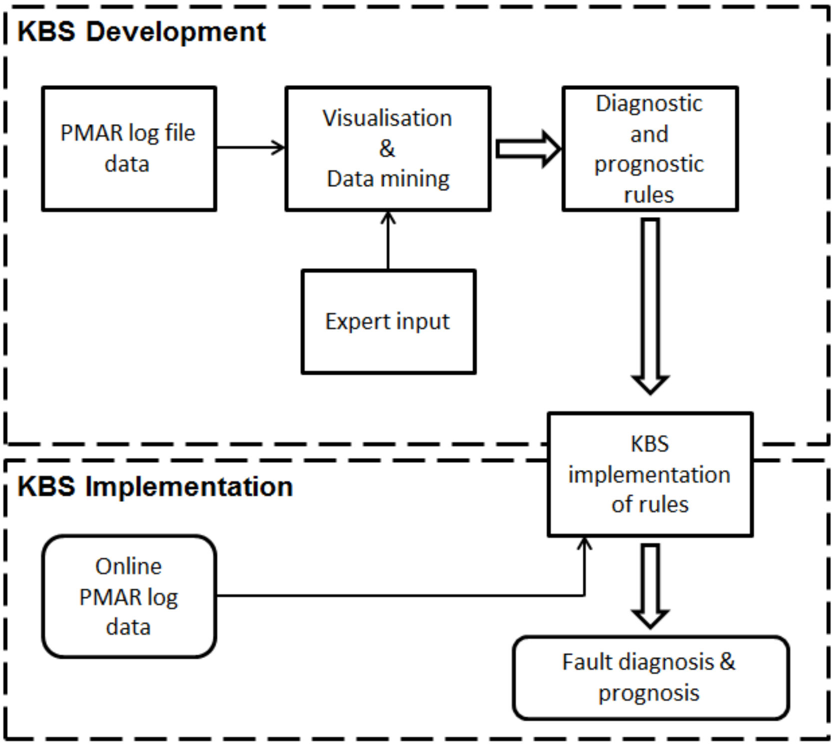


Figure 4.6: Development and implementation of KBS

and definitions of distribution network operations, which are mostly written in the form of rules. Therefore, for the purpose of keeping consistency and flexibility of the designed DSS, the deployed KBS concentrates on utilising RB strategy for automatic data analysis by moving ready-made rules and generating practical rules through knowledge engineering and data science techniques. Figure 4.6 sketches the overall KBS development process based on utilising data science and expert knowledge, followed by its implementation within the DSS.

As shown in Figure 4.6, the development of the KBS is based on deriving and defining the diagnostic and prognostic rules, which are generated through visualisation and data mining of actual PMAR historical data, and is also supported by expert knowledge and experience. In order to implement the functionality, a KBS is used to deploy the fault analysis rule-base to process the online PMAR data. It invokes the appropriate rules based on the input data, for both the diagnostic and prognostic functions [WMS+17].

4.3.3.1 RB Reasoning Strategy and Development Platform

Since the KBS adopts the RB strategy, the primary task of developing the KBS is to choose an appropriate reasoning strategy (i.e. forward-chaining, backward-chaining, or a combination of the two) [Lug09]. This directly affects the determinations of the platform and associated RB inference engine being applied for the development of the KBS.

Concerning the fault diagnosis and prognosis is based on the analysis of PMAR data, the data associated with the conditions of PMAR devices or circuits are prepared and known, the objective is to conclude the diagnostic or prognostic results and required actions based on the provided data. This situation is naturally data-driven, so the forward-chaining approach is effectively suitable for developing the KBS. In addition, the forward-chaining strategy is also suited to the automated manual process, which searches the relevant information with matched conditions. As a result, the forward-chaining strategy is adopted in this work.

There are a number of available rule inference engines which help to process rules efficiently. Two of the most common rule engines are Drools and Java Expert System Shell (JESS) [Red13]. This type of rule engine provides the developers with a platform to inject and convert a set of knowledge into a sequence of rules. The differences between these two rule engines are the different formats for writing rules and reasoning rules. In this thesis, Drools has been selected as the platform for implementing the automatic KBS. It provides a forward-chaining reasoning platform which develops and manages the diagnostic and prognostic rules, with a set of benefits described as follows:

- Easier to translate and understand rules: compared with procedural code, Drools rules are much easier for business analysts or new developers to understand the solutions of difficult problems. It also provides explanations about the adopted solutions and final decision.

- High flexibility and inference efficiency: Drools separates logic and data, this makes the management of rules easier that other inference engines, Drools uses Rete algorithm [For82] to provide efficient ways of matching appropriate rule pattern to the objects (i.e. data of PMAR device's or circuit's conditions).

- Reasonable performance and ease of development: the operational environment of Drools is Java, which offers good compatibility when interfacing with any other operating systems. This prevents the developers from being constrained by particular systems or hardware platforms when designing the logic system. Furthermore, Drools can be easily integrated with development tools, such as Eclipse [Ecl14] (and in future, the web based UIs). These tools can provide ways to efficiently manage, audit, and debug the rules to obtain quick feedback and validation.

Within Drools, the rules will be represented in the form of "when-then", where the "when" defines the conditions for matching the rule and the "then" indicates the actions after firing the rule. As mentioned in the previous text, Rete algorithm is responsible for processing the reasoning of rules, which separates the matching algorithm into two types based on the problems' patterns:

- One-Pattern: means the problem has a single matching path or pattern to reach the rule execution.
- Multi-Patterns: is where the problem has multiple pattern matches, that means when a list of conditions are fulfilled in the rule, the algorithm will then give the execution.

Figure 4.7 shows the structure of a Drools rule, where the type of problem is multi-patterns. Within the rule, 'attributes' are also defined to control the rule execution before the Rete algorithm tries to match the object with the listed conditions. The Drools rules are written in simple text-based files, which can be easily accessed and updated.

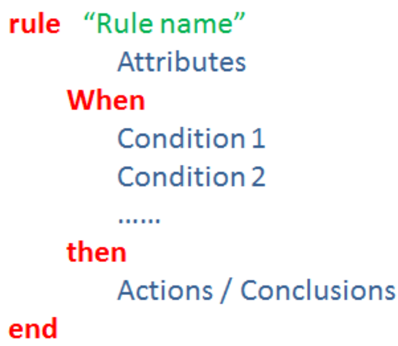


Figure 4.7: The structure of a Drools rule

4.3.3.2 The Structure of KBS

The inference engine acts as a key part of designing the KBS, Figure 4.8 illustrates the overall structure of the KBS, including the inference engine, production memory and working memory. The production memory is used to store the diagnostic and prognostic rules and to perform data analysis. The working memory is used to store the input data (i.e. the facts required to be reasoned) for temporary inference. In this thesis, the rule base contains the rules for automatic fault diagnosis and prognosis, and depending on the different functions, these rules are stored separately in the dedicated rule package. This can ensure only the associated rules will be invoked without being erroneously affected by irrelevant rules when the inference engine performs a reasoning task.

At the centre of the structure is the Drools inference engine, which contains the pattern matcher and the agenda. The function of pattern matcher is to match the facts from working memory to the relevant rules in the production memory based on the defined conditions. When all the conditions are matched within a particular rule, the conclusion of the rule will be executed. If multiple rules are matched simultaneously, the agenda will determine the priority of executing the rules to prevent the conflicts between rules, this process can be achieved by defining the salience values in the attributes of rules [Red13].

In order to achieve the function of data interaction between the production memory and the working memory with the inference engine, two elements called rule selector and knowledge session are used in the data processing of KBS. When

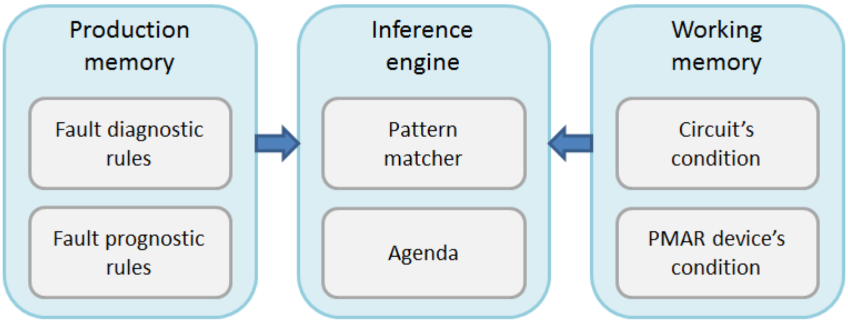


Figure 4.8: The structure of KBS

the KBS is performing a reasoning task of fault diagnosis and prognosis, the rule selector will retrieve the correct rule files based on provided case-specific information. For instance, when the KBS needs to analyse the data for diagnosing PMAR device faults, the rule selector is required to consider the information to choose the appropriate rules by validating the PMAR's name, condition, available recorded messages, etc. After capturing the right rules, the KBS will create a new knowledge session, which obtains the relevant facts (i.e. PMAR data) and combines them with the selected rule to interact with the inference engine for reasoning. Such processing in the Drool engine allows multiple simultaneous tasks by creating different knowledge sessions. That means that the KBS could analyse the data from multiple PMARs at the same time. Although the KBS performs different knowledge sessions at the same time, each of the sessions are independent without affecting the others. Figure 4.9 demonstrates the relationship between the knowledge session and the inference engine when KBS is performing a task.

4.3.3.3 Reasoning Process of Fault Diagnosis and Prognosis

With respect to the descriptions of the structure of KBS and its internal inference engine and data interaction, Figure 4.10 shows the reasoning process of fault diagnosis and prognosis of the KBS which follows the automated process of identifying the problematic PMAR devices. The main processing steps are:

1. The information of the identified problematic PMAR devices will be inserted into the KBS, to select the related PMAR log files for starting the automatic fault diagnosis and prognosis.

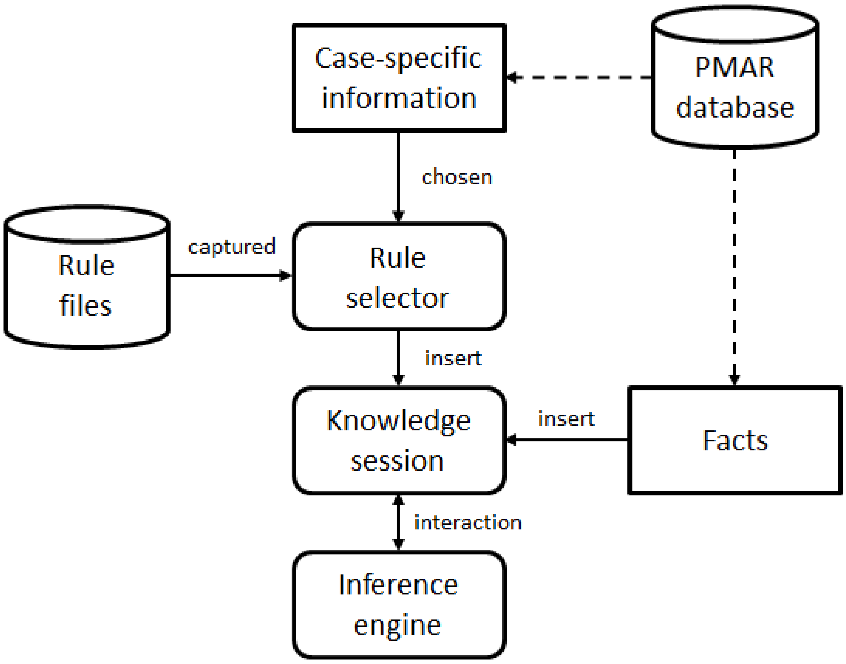


Figure 4.9: Data interaction with knowledge session and rule selector

2. The PMAR log data (contains recorded circuit behaviour and device's condition), the case-specific information, and the associated rule files are captured by the rule selector for creating a new knowledge session for data analysis.

3. A knowledge session is created by the KBS, which interacts with the inference engine for diagnostic and prognostic tasks.

4. The inference engine will process the imported data and match it with the associated rules based on the requirement of the automatic data analysis. These rules are divided into: sections of diagnosis of PMAR device fault, identification of a semi-permanent fault, and the prediction of potential PMAR operations.

5. The diagnostic and prognostic results will be fed into the GUI for assisting engineers in the decision-making process, and some of them will be transferred to the visualisation tool for further analysis.

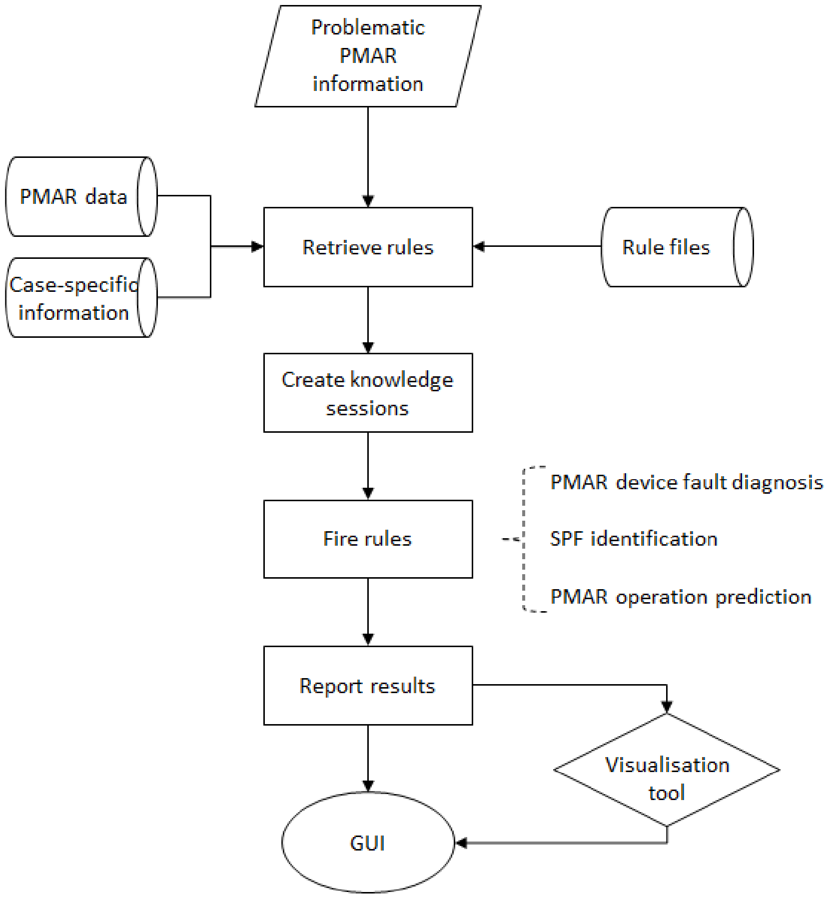


Figure 4.10: The process of KBS for fault diagnosis and prognosis

4.4 Conclusion

This chapter detailed the design of a novel DSS for automatic fault diagnosis and prognosis by developing a KBS analysing PMAR data. This system converts the daily manual analysis of SCADA alarm data into an automated process, and continues to diagnose the underlying PMAR devices' and circuits' conditions with related auto-recloser data. This application of data analytics is supported by the implementation of a KBS to achieve reactive fault management and post-fault investigations into anomalous network behaviour. Meanwhile, the DSS 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 potential outages.

For the development of the KBS, this work applied the RB strategy to diagnose the nature of a circuit's or PMAR device's fault activity and highlights indications of problematic events that might gradually evolve into a full-scale circuit fault. These processes will be achieved by invoking the rules for: diagnosing PMAR device fault, detection of semi-permanent faults, and prediction of potential PMAR operation. All the data analysis steps which are integrated with identifying the problematic PMAR devices will be automated. In this chapter, the KBS was introduced with its design and development. The implementation of the KBS will be detailed in Chapter 5 and 6 with generating the diagnostic and prognostic rules from expert knowledge, data mining, and clustering techniques. Each of the chapters will demonstrate case studies for validating the functionalities of automatic fault diagnosis and prognosis designed in the KB DSS.

Chapter 5

Automatic Fault Diagnosis Functionality with Case Studies

5.1 Introduction

As described in Chapter 4, this thesis introduced the design of the decision support system to assist engineers' data analysis and management of distribution network operations and control. The developed DSS deploys a KBS to achieve the proposed diagnostic and prognostic functionalities, which utilises the RB strategy to process the automatic analysis. In this chapter, the implementation of the fault diagnostic aspect of the KBS is presented along with details of the development of the associated rules. These are demonstrated and consolidated with appropriate case studies.

This chapter begins with an introduction of what knowledge is considered for PMAR device fault diagnosis and identification of semi-permanent faults and how the knowledge is used for generating diagnostic rules. Then, the implementation of fault analysis rules is partitioned into two parts which are based on the two most important diagnostic functionalities. In order to demonstrate the performance of the developed rules, the thesis will briefly introduce the prototype of the designed DSS before utilising the integrated visualisation tool and GUI to illustrate the mechanism of the fault diagnostic rules through case studies.

5.2 Knowledge for Fault Diagnosis

As highlighted in the previous chapter, the knowledge-based system serves as the most suitable technique to solve the problems confronted in this work, which possess a number of advantages. The main advantage of applying a KBS is the diagnosis of the PMAR device faults and monitoring of the circuit behaviours can be directly supported by utilising expert knowledge and network operation policies including fault or abnormal event explanation and justification. For the purpose of extracting diagnostic information and transforming relevant knowledge to develop the rule base for implementing the KBS, the knowledge engineering (involving knowledge elicitation, interpretation, validation and utilisation) [SBF11] becomes the most important processing stage in developing a comprehensive diagnostic rule base. The details of each processing step of knowledge engineering and related tools can be found in [SWH94, SAA+00]. This section will introduce the types of knowledge aggregated and the subsequent forms of elicitation which are required to develop a rule base for automatic fault diagnosis.

Concerning the definition of knowledge described in section 3.4, the knowledge can be simply divided into explicit knowledge and tacit knowledge. The explicit knowledge can be easily defined and understood, but tacit knowledge is obtained by domain experts through the persistent application and learned practical experience to obtain proficiency within a subject area and can be difficult to quantify. For instance, the knowledge about driving a car is gained through personal training rather than reading the handbook. In this research, both of these two types of knowledge are adopted for the knowledge elicitation from the experts to construct the KBS. Following the elicitation, the processing of knowledge interpretation, validation and utilisation will be presented in the next sections.

Generally speaking, the knowledge used for designing and developing the diagnostic rule base mainly consists of domain experts' understanding and experience, system equipment' settings and network operation policies. Furthermore, the appropriate knowledge was gathered through regular interviews and dialogue between the experts in the field of distribution network operation and protection (who work with the) the DA devices, i.e. PMARs) and the researcher. Each interview is scheduled by an agenda meeting involved with a presentation part and a discussion part to record current research and future work. Throughout these meetings, valuable information and understanding of domain knowledge was captured and utilised to build the rule-base for automatic fault diagnosis. The overall process of knowledge translation and management is shown in Figure 5.1.

As demonstrated in the figure, The knowledge translation for fault diagnosis (i.e. knowledge elicitation and knowledge interpretation) is on the basis of generic knowledge, experts' understanding and experience when evaluating circuit behaviours and asset conditions in the distribution network. The generic knowledge includes settings and network policies. Here, the settings specifically refer to the preferences of the deployed PMARs, which satisfy the requirements of network operation and protection as the priority. The PMARs' settings can be adjusted by the engineers to offer more flexibility to control the devices in the DA schemes as well. For example, the control engineers can determine the appropriate frequency of data transportation between the PMARs and the SCADA system hubs to ensure the control centre captures applicable data for power system analysis. For the network operation policies, these are a set of official documents which define the regulations for reliable and stable electricity service relating to how the protective PMARs should be configured, and specify the criteria that the equipment' settings must conform to. For example, in areas with poor weather conditions, it accelerates the deterioration of installed devices which can result in more frequent transient faults or longer duration of individual transient fault on the overhead lines. In order to prevent long-term outages occurring after the lockouts are tripped by the PMARs, the number of auto-reclosing attempts (the PMAR device remains open for a period of 10 seconds before attempting a reclose) for clearing the fault can be increased from 2 to 3 before the registration of a lockout. That means the PMAR will 'lockout' on its fourth attempt, which allows another 10 seconds for the PMAR to automatically clear the fault.

To suppress this issue without the underlying requirement of equipment spe-

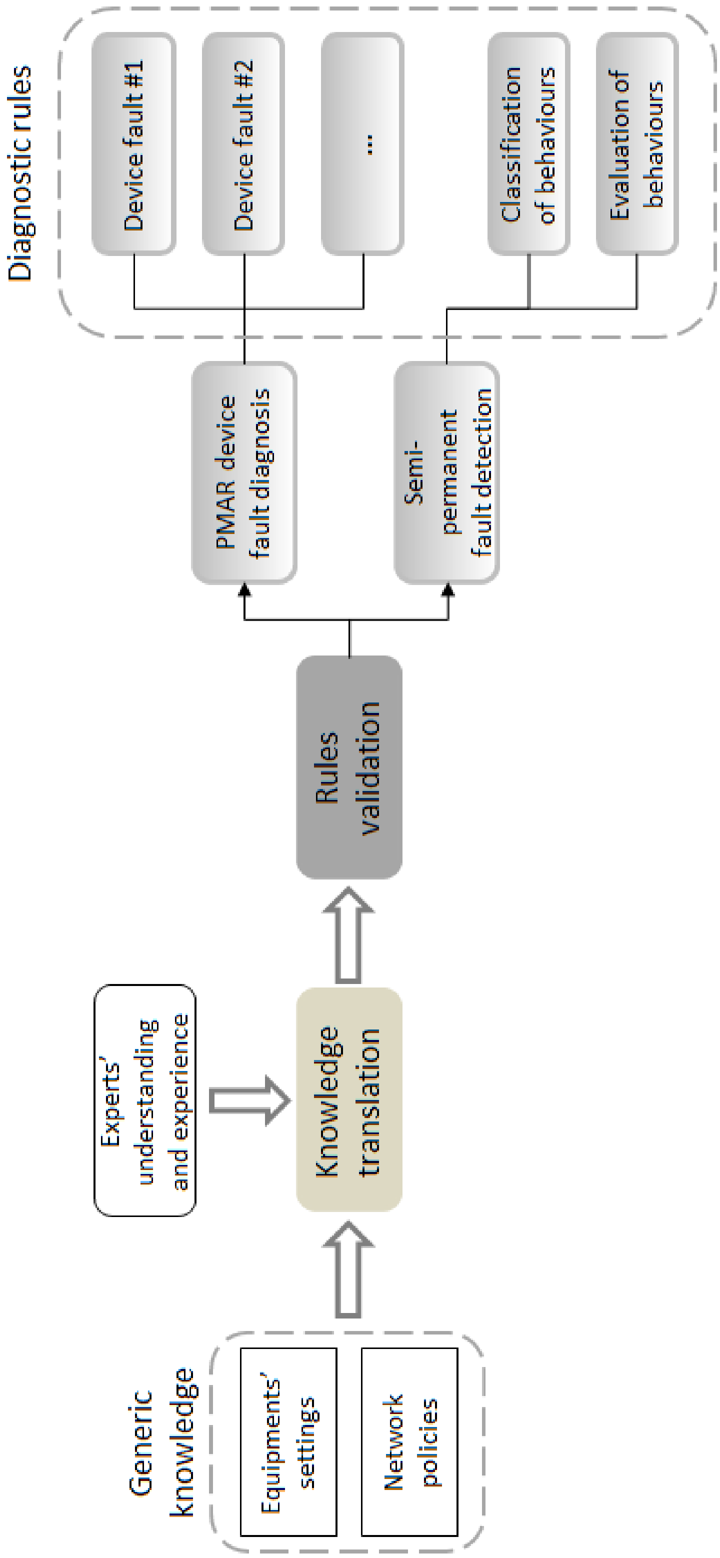


Figure 5.1: Knowledge translation and management

cific parameters and network policies, expert knowledge and experience of fault diagnosis can be gathered and presented in various ways. The knowledge could be either an extended definition of the network's settings and policies, or the summarised experience of operating and protecting the circuits. For instance, the definitions of classifying transient, semi-permanent, permanent faults are one type of experts' knowledge based on the experience of analysing the circuit behaviours. Furthermore, no matter what types of experts' knowledge is gathered from engineers and translated to diagnostic rules, the knowledge should be validated to ensure that it accurately reflects the experts' understanding and experience of the previous confronted scenarios. A process for validating the rules transformed from the facilities' settings and network policies is also required to guarantee the rules align with the associated expert knowledge. These validation activities can be evaluated through validation meetings with the relevant experts to review the knowledge elicitation and translation and ensure that there is no misinterpretation of the knowledge that is being captured. Meanwhile, tests of the defined rules can be undertaken to determine whether the rules can be suitable for various situations.

After the knowledge has been validated, any changes to the knowledge will be incorporated within the rules to be utilised in the KBS. As shown in Figure 5.1, both of the setting policies and experts' knowledge are incorporated in the process of developing the rules, which are categorised into two separate groups to diagnose the PMAR device faults and detect semi-permanent faults on the circuits respectively. For identifying the PMAR device faults, each defined rule can be operated independently to check the expected device fault, because the fault itself has specific features. However, for identifying the semi-permanent faults, the detection process can be segregated further into two additional stages, including the first step to categorise the circuit behaviour based on different PMAR operations confronting with faults, and then to evaluate the trends of defined features to detect semi-permanent faults. Within the case studies in Chapter 5.6, the automatic fault diagnosis will be demonstrated and subsequently supported by visualisation tools. Examples of the rules will be provided in the following sections to convey how the knowledge translation and rule development are achieved.

5.3 Diagnosis of PMAR Device Faults

PMARs act as an approved type of protective devices widely installed on the SPEN distribution system, which provide significant protection against fault and abnormal operating conditions. For the purpose of supplying reliable services, the physical health of the device is one of the key factors to ensure successful operation. Therefore, engineers need to check the status of the relevant PMAR device regularly. However, this process can be inefficient since this would require engineers to be dispatched to the physical location of the relevant devices where they would then check and evaluate said device before reporting back. Although the time-consuming process could identify parts of device faults by observing warning flash-lights corresponding to particular permanent faults and detecting the alarm messages displayed in the PMAR's control panel, abnormal behaviour or operational degradation may not be detected or registered on-site due to underlying device faults or human error. These anomalies could lead to failure operation of the PMAR device and potentially result in permanent device faults impacting online services in future. As a result, to adequately diagnose the PMAR device faults for providing fast and accurate assessment reports, it is important to automatically analyse the underlying information on the PMARs' conditions. Fortunately, both of the indicative alarms and anomalous information can be retrieved from the Main Processor Module (MPM) integrated with the inside control panel.

As introduced in previous subsection 2.4.3, the MPM is the core component of PMAR for condition monitoring, which also controls the operation of the PMAR. The MPM records details (sampled at 12.8kHz, detailed in Appendix A) of online activity (i.e. fault current and voltage amplitudes) and the condition of the PMARs components. So, depending on the implemented rule-base, the KBS can automatically detect them (based on the knowledge of operational engineers and setting policies), which characterise common known PMAR faults. The aim is to prevent potentially delayed or failed operations in response to overhead line faults in addition to analysing the undetected damages with PMAR's components. These diagnostic rules operate on the automatically imported MPM's records represented in the PMAR log files, detailed in subsection 2.4.3. These faults are identified through: interpretation of alarms generated by the MPM; calculations on interval times between the PMAR operations and status changes within the log file. This section shows some examples of the development of diagnostic rules for identifying PMAR device faults.

5.3.1 PMAR MPM Fault Diagnosis Rule

The MPM fault indicates the whole PMAR device is not operating as expected and therefore defective. There are two main causes which primarily lead to the MPM faults, i.e. the permanent damage of the integrated circuits within the processor module and the MPM firmware version being out of date. This will then result in device failure for an in-zone fault. The MPM fault is the only type of the device faults that can be both identified through checking control panel manually and invoking the rule with imported log file automatically. The other PMAR device faults can only be diagnosed by the KBS through analysis of the data. Although the operating engineers can check the MPM fault by viewing the displayed multiple flashing LEDs, the indicative alarms may also fail simultaneously because of the damage to the circuit board. Therefore, the most reliable method is to interrogate the PMAR log files to perform a comprehensive data analysis, which can be achieved automatically through the decision support system.

The approach for confirming an MPM fault is to simply detect the "MPM fault" alarms in the log file, which were generated automatically by the MPM. However, under some special situations, such as the replacement of panel's battery or the upgrade of the MPM firmware version, several "MPM fault" messages could be mistakenly generated during a short period in the log file. To prevent the KBS from reporting false diagnostic results, the developed rule should select the 'true' perations / fault records | Change messages | Load profile (+) | Load profile (-) | Event log | Unsolicited responses

Date and Time	Event title	Start/End	Source of event	Relevant phase
19/04/2013 17:56:12:591	MPM fault	End	ISC	
19/04/2013 18:00:32:739	MPM fault	Start	ISC	
19/04/2013 18:02:42:904	MPM fault	End	ISC	
19/04/2013 18:07:03:915	MPM fault	Start	ISC	
19/04/2013 18:09:14:132	MPM fault	Start	ISC	
19/04/2013 18:11:24:263	MPM fault	End	ISC	
19/04/2013 18:22:15:313	MPM fault	Start	ISC	
19/04/2013 18:24:25:638	MPM fault	End	ISC	
19/04/2013 18:26:36:085	MPM fault	Start	ISC	
19/04/2013 18:28:46:533	MPM fault	End	ISC	
19/04/2013 18:30:56:802	MPM fault	Start	ISC	
19/04/2013 18:33:07:059	MPM fault	End	ISC	
19/04/2013 18:50:29:236	MPM fault	Start	ISC	
19/04/2013 18:52:39:404	MPM fault	End	ISC	
19/04/2013 18:56:59:645	MPM fault	Start	ISC	
19/04/2013 18:59:09:948	MPM fault	End	ISC	
19/04/2013 19:07:50:969	MPM fault	Start	ISC	
19/04/2013 19:10:01:369	MPM fault	Start	ISC	
19/04/2013 19:12:11:765	MPM fault	End	ISC	
19/04/2013 19:31:43:848	MPM fault	Start	ISC	
19/04/2013 19:33:54:117	MPM fault	Start	ISC	
19/04/2013 19:36:04:365	MPM fault	End	ISC	

Figure 5.2: A typical MPM fault with fleeting fault messages

alarm messages. Generally, when an MPM fault occurs, it will last for a few hours and the MPM will generate around 10 alarms per hour in the log file, as shown in Figure 5.2.

Hence, based on the nature of an MPM fault and experts' experience of confronting the MPM fault issues, a critical amount (i.e. 20 consecutive alarms was set as the threshold with domain expert's experience) becomes the criterion of invoking the diagnostic rule. Figure 5.3 shows the rule that determines whether an MPM fault exists in the imported log file. The rule initially searches for each "MPM fault" event in the log file, and if an "MPM fault" is found and it exceeds the critical alarm count, it implies that the specific PMAR has been affected by an MPM fault. Then, the details of this fault will be reported in the form of text messages displayed in the designed user interface. The corresponding source code is shown in Figure 5.4 to provide an example of the operation of Drool's rules.

```
rule      "Identify MPM fault"
    when
            Consecutive "MPM fault" events are detected
            And
            The number of events is larger or equal to 20
    then
            Report the result and message
end
```

Figure 5.3: Rule for identifying the MPM fault

Search for an "MPM fault" event

```
rule "Main Processor Module Fault"
    when
        $mpmfilter : device( col1 == "MPM fault", No : indexNo);
        exists device( col1 == "MPM fault", indexNo == No+9 );
        exists device( col1 == "MPM fault", indexNo == No+8 );
        exists device( col1 == "MPM fault", indexNo == No+7 );
        exists device( col1 == "MPM fault", indexNo == No+6 );
        exists device( col1 == "MPM fault", indexNo == No+5 );
        exists device( col1 == "MPM fault", indexNo == No+4 );
        exists device( col1 == "MPM fault", indexNo == No+3 );
        exists device( col1 == "MPM fault", indexNo == No+2 );
        exists device( col1 == "MPM fault", indexNo == No+1 );
        exists device( col1 == "MPM fault", indexNo == No-1 );
        exists device( col1 == "MPM fault", indexNo == No-2 );
        exists device( col1 == "MPM fault", indexNo == No-3 );
        exists device( col1 == "MPM fault", indexNo == No-4 );
        exists device( col1 == "MPM fault", indexNo == No-5 );
        exists device( col1 == "MPM fault", indexNo == No-6 );
        exists device( col1 == "MPM fault", indexNo == No-7 );
        exists device( col1 == "MPM fault", indexNo == No-8 );
        exists device( col1 == "MPM fault", indexNo == No-9 );
        exists device( col1 == "MPM fault", indexNo == No-10 );
    then
    $mpmfilter.setSign1(1);
        msg1.append("Main Processor Module Fault" + No + "---"+ $mpmfilter.getSign1());
        msg2.append("A number of fleeting MPM alarms");
end
```

Condition: if number of detected consecutive "MPM fault" alarms is larger or equal to 20

Prepare the message to display the final result with an explanation

Figure 5.4: The source code for identifying an MPM fault

5.3.2 PMAR Driver Module Fault Diagnosis Rule

The driver module fault diagnosis rule is provided to interrogate the status of the contacts and switches with the PMAR, i.e. the rule checks whether the driver module performs as required from setting policies. The driver module acts as an intermediary, which receives the 'Trip/Close' control signals from the MPM and converts them into current pulses applied to the magnetic actuator coil to drive the contacts into the open or closed position. It also converts the PMAR's auxiliary switch status into a logical position signal for use by protection and indication elements of the MPM. Therefore, the health of the PMAR's coil circuit and the driver's own readiness to execute the next 'Trip/Close' operation is monitored by the driver module.

Usually, when a driver module fault occurs, it can be represented by two modes: the PMAR failed to operate for a fault within protection zone, or the PMAR did not complete the auto-reclosing sequence. Either fault will generate warning events and therefore be logged by the MPM into the log file. These recorded warning messages include "Driver not ready", "OSM coil SC", "OSM coil isolated", "Excessive To", and "Excessive Tc". Each of them indicates a driver module fault with particular possible cause. The "Driver not ready" signal indicates the driver cannot execute the next control operation ('Trip/Close') because the capacitors of the driver are not sufficiently charged. The "OSM coil SC" or "OSM coil isolated" signals the potential damage of the coil resulting in the problems with driver's operations. The "Excessive To" (i.e. contact opening time exceeds setting time) and "Excessive Tc" (i.e. contact closing time exceeds setting time) means the driver does not complete the auto-reclosing sequence in the set duty cycle. So, the rule in the KBS to check a driver module fault is based on the detection of specific alarms, as shown in Figure 5.5.

5.3.3 PMAR Tank Fault Diagnosis Rule

The PMAR tank fault diagnosis rule is responsible for identifying whether an issue with the tank will lead to anomalous or failure operation of the PMAR. In

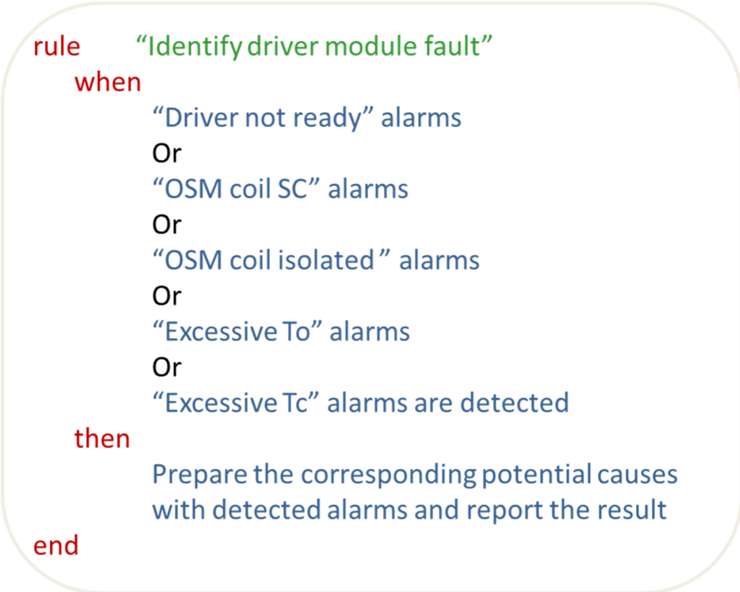


Figure 5.5: Rule for identifying the driver module fault

general, a tank is constructed from stainless steel with a rated long lifetime under normal circumstances. Unlike other parts of the PMAR, the tank is significantly more robust to environmental conditions and potential damage from livestock. If a common tank fault is detected, it is usually permanent damage, which may be a result of deterioration due to ageing, or bad weather (e.g. lightning strikes). Under these expected fault scenarios, the conclusion is to suggested a change of tank.

With a tank fault, the magnetic actuator will not apply the parallel connected auxiliary switches to the default positions of the mechanism correctly. This means that the recommended values for the open and close operations can be unsuitable. The MPM will repeatedly generate “Open/Closed UNDEFINED” events with unrealistic time scales (for the realistic time scale, the time duration of changing status from ‘Open’ to ‘Closed’ should be slightly larger than 10 seconds) into the log file, which is shown in Figure 5.6. Depending on the typical cases of tank faults, a threshold is set by the engineers to judge the amount of repeat logged events for determining the existence of a tank fault, and this is visualised alongside the diagnostic rule in Figure 5.7.

perations / fault records | Change messages | Load profile (+) | Load profile (-) | Event log | Unsolicited res

Date and Time	Event title	Source of event	Relevant state	Critical parameter
Data may be lost				
10/03/2012 08:38:54:758	Open	Manual	Lockout	
10/03/2012 08:38:54:923	Closed	Undef		
10/03/2012 08:38:55:569	Open	Manual	Lockout	
10/03/2012 08:38:55:584	Closed	Undef		
10/03/2012 08:38:55:764	Open	Manual	Lockout	
10/03/2012 08:38:55:784	Closed	Undef		
10/03/2012 08:38:55:809	Open	Manual	Lockout	
10/03/2012 08:38:55:949	Closed	Undef		
10/03/2012 08:38:57:071	Open	Manual	Lockout	
10/03/2012 08:38:57:352	Closed	Undef		
10/03/2012 08:38:57:898	Open	Manual	Lockout	
10/03/2012 08:38:57:983	Closed	Undef		
10/03/2012 08:38:58:509	Open	Manual	Lockout	
10/03/2012 08:38:58:604	Closed	Undef		
10/03/2012 08:38:58:699	Open	Manual	Lockout	

Figure 5.6: A typical tank fault with repeat undefined events

```
rule    "Identify tank fault"
  when
        Consecutive driver status changes ('Open' to 'Closed'
        time is less than 10 seconds)
        And
        The number of events is larger or equal to 10
  then
        Report the result and message
end
```

Figure 5.7: Rule for identifying the tank fault

5.3.4 PMAR Microswitch Fault Diagnosis Rule

The PMAR reports its position status to the auto-recloser control cubicle using microswitches. Their status is opposite to the main contacts and can be checked on the control panel. So if the PMAR is open, the microswitch is closed. On the contrary, when the auto-recloser is closed, the status of microswitch should be open. Therefore, the microswitch fault diagnosis rule contains the knowledge that defines the criteria to determine whether a fault incorrectly reflects the main contacts' position status. Under normal conditions, when the MPM signals the driver module to execute an action (Trip/Close), the contact needs a reaction time, and the reaction time should be within a pre-determined time range. If the time duration between the contact receives the signal and auto-recloser status changes exceeds the default time range, then a microswitch fault is indicated. The PMAR product manual states that contact closing time (i.e. the duration from signal 'Close' to status position of 'Closed') exceeding 100ms or contact opening time (i.e. the duration from signal 'Trip' to status position of 'Open') exceeding 60ms can be treated as excessive operating times. Furthermore, the time out status is indicative of a microswitch fault.

Figure 5.8 shows an example of correct PMAR operation without an indication of a microswitch fault. The highlighted correct times is the standard duration between the open and close operations (i.e. 10 seconds between open and close operations is a function of the protection system set by the network policies). In Figure 5.9, an example of an auto-reclosing sequence from a PMAR with a suspected microswitch issue is presented. The sequence is incomplete because the time difference between open and close operations is too small and the auto-recloser, recognising the abnormal condition, locks into open position.

Meanwhile, if the time between 'Trip' signal and status 'Open' indication is less than 25ms and time between 'Close' signal and status 'Closed' indication is less than 40ms, this can also suggest the microswitch may be faulty. Hence, the KBS identifies this fault by comparing the actual time periods of status changes with primary settings through automatic time stamp calculations; the

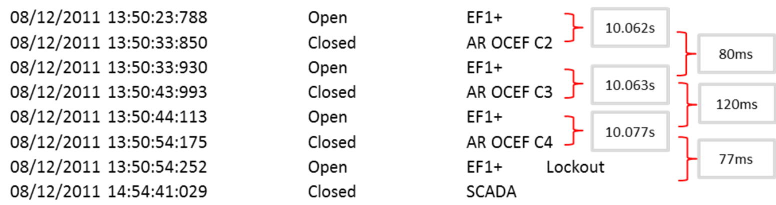


Figure 5.8: A typical correct PMAR operation without a microswitch fault

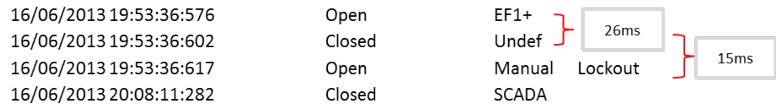


Figure 5.9: A typical PMAR operation with a suspected microswitch fault

corresponding rule is shown in Figure 5.10.

5.3.5 PMAR Umbilical Cable Fault Diagnosis Rule

The umbilical cable (i.e. control cable) has a male and female connector, the female connector fits into the recloser control cubicle, and the male connector fits into the base of the recloser. All the log data and signals are transmitted through the umbilical cable to ensure the correct PMAR operation and control. Therefore, the condition monitoring of an umbilical cable is critical in the health check of PMAR's components. By detecting the damage of the umbilical cable, it could prevent potential significant issues on operations of the PMAR in future. Due to the umbilical cable always being exposed to the natural environment in addition to the plastic shielding of the cable, it often suffers deterioration. In most cases, the ageing umbilical cable will suffer from water ingress, and if this occurs the umbilical cable should be replaced.

The umbilical cable fault diagnosis rule for identifying the fault is to examine the log file with specific fleeting alarms. The registered alarms are marked with "Pickup" event from the source of "Uabc>", where the "Uabc>" indicates the maximum voltage is activated on the three-phases (i.e. A, B and C terminals), that is caused by the water ingress. Figure 5.11 shows the fleeting alarms caused by water ingress to the umbilical cable retrieved from a typical log file. For developing the diagnostic rule, the knowledge sets an invoking threshold (was set

```
rule      "Identify microswitch fault"
    when
          Time interval between Action 'Trip' to Status 'Open' is more than 60ms
          Or
          Time interval between Action 'Close' to Status 'Closed' is more than 100ms
          Or
          (
          Time interval between Action 'Trip' to Status 'Open' is less than 25ms
          And
          Time interval between Action 'Close' to Status 'Closed' is less than 40ms
          )
    then
          Report the result and message
end
```

Figure 5.10: Rule for identifying the microswitch fault

with consideration of domain expert's experience) based on the characteristics of an umbilical cable fault, which is displayed in Figure 5.12.

5.4 Diagnosis of OHL Distribution Circuit Degradation - Detection of Semi-Permanent Fault

As presented in section 5.3, the first stage of automated decision support system focuses on PMAR device fault diagnostics. In this section, the KB DSS concentrates on the diagnosis of overhead line faults. As discussed previously, depending on the performances and reactions of PMARs, the overhead line faults can be divided into three categories: transient faults, semi-permanent faults (SPFs) and permanent faults. Among these types of faults, the SPF is the most difficult type to detect and confirm, but also the most valuable type to identify for increasing the reliability and security of the power system because successful detection of this type of faults will allow preventive actions to be performed to avoid irreparable damage. Therefore, the detection of the SPF will be the key focus of the work of the diagnosis of overhead line faults.

For detecting the SPF, the designed system breaks down the process into two rations / fault records | Change messages | Load profile (+) | Load profile (-) | Event log | Unsolicited resp

	Date and Time	Event title	Start/End	Source of event	Relevan
▸	27/03/2013 07:49:49:293	Pickup	Start	Uabc>	
	27/03/2013 07:58:43:899	Pickup	End	Uabc>	
	27/03/2013 07:58:44:160	Pickup	Start	Uabc>	
	27/03/2013 08:22:54:678	Pickup	End	Uabc>	
	27/03/2013 08:22:55:476	Pickup	Start	Uabc>	
	27/03/2013 08:29:08:066	Pickup	End	Uabc>	
	27/03/2013 08:29:08:246	Pickup	Start	Uabc>	
	27/03/2013 08:29:08:267	Pickup	End	Uabc>	
	27/03/2013 08:29:08:587	Pickup	Start	Uabc>	
	27/03/2013 08:31:19:590	Pickup	End	Uabc>	
	27/03/2013 08:31:19:910	Pickup	Start	Uabc>	
	27/03/2013 08:35:40:883	Pickup	End	Uabc>	
	27/03/2013 08:35:41:581	Pickup	Start	Uabc>	
	27/03/2013 08:44:18:370	Pickup	End	Uabc>	
	27/03/2013 08:44:18:530	Pickup	Start	Uabc>	
	27/03/2013 08:49:42:115	Pickup	End	Uabc>	
	27/03/2013 08:49:48:643	Pickup	Start	Uabc>	
	27/03/2013 08:56:27:711	Pickup	End	Uabc>	
	27/03/2013 08:56:33:178	Pickup	Start	Uabc>	
	27/03/2013 08:56:51:238	Pickup	End	Uabc>	

Figure 5.11: The log file shows fleeting alarms caused by water ingress to the umbilical cable

```
rule     "Identify umbilical cable fault"
   when
         Consecutive "Pickup" events with the source of "Uabc>" are detected
         And
         The number of events is larger or equal to 10
   then
         Report the result and message
end
```

Figure 5.12: Rule for identifying the umbilical cable fault

coherent stages to strip out the SPFs from thousands of fault events recorded in the log files, based on the automatic data analysis and experts' decisions. The two processing steps: classification of PMAR operations and evaluation of behavioural trends for SPF detection, which are briefly introduced in the section of knowledge translation and management. As a SPF manifests itself through sporadic periods of intense PMAR operation on a circuit, it is therefore necessary to first classify the circuit behaviour into different classifications of PMAR operation. The first stage (as shown in Figure 4.4)is data acquisition where we capture the events within the PMAR log files, and then, the next step is to filter the activity consistent with SPFs. The suspicious SPF activities will be considered through trend evaluation to confirm the existence of SPFs [WMS+17].

To assist the detection of potential SPFs, a further rule-base was developed to automatically classify the PMAR operations being experienced. The rules were developed using expert knowledge from engineers and knowledge of the PMARs operating mechanism. Details of a PMAR operating mechanism can be found in section 2.4.3. Following the classification, a visualisation tool built within the decision support system is used to allow engineers to observe the results and allows them to determine potential SPF activities through identifying data trends and patterns associated with PMAR operations. In addition, data can also be visualised to substantiate the diagnosis of invoked rules. The following subsections describe how the diagnostic rules are generated and the trends evaluated to detect SPFs. The visualisation tools are described in the following section 5.5.

5.4.1 Classification of PMAR Operations

As described previously in subsection 4.2.2, the PMAR operations depend on the fault current amplitude and its duration, which is recorded as 'pick-up' activity in the PMAR log file [WSKM14]. When the MPM registers a pick-up activity due to excessive current amplitude, a trip signal to open the contact of the auto-recloser will be considered. If the pick-up has a very small duration, the PMAR will not take any actions; but if the anomaly lives over the time threshold of executing a trip, the PMAR will provide one or several reclosure attempts (the number is set by the operator to the specific network) to clear the fault. However, if the fault remains on the overhead line after all default reclosure attempts, a ‘lockout’ will be triggered to protect the circuit from damage, where the lockout permanently opens the faulty area, and can only be reclosed manually or via telecontrol by the operational engineers.

Based on the fundamental knowledge of PMAR operations, using the log file pick-up data recorded during the operation of PMARs under different fault conditions, a set of specific rules are defined to classify the PMAR operations into different classifications for grouping the anomalous activities consistent with SPF candidates. These rules are based on the experts’ understanding and experience of how these faults are likely to manifest themselves in these data sets. Therefore, the rules divide the fault pick-up activities into four classifications: resulting in no trip (FP); single trip (ST); multiple trip (MT); and, lockout (L). Figure 5.13 shows the rules for classifying PMAR operations (i.e. FP, ST, MT and L) in order of increasing severity.

As displayed in the figure, the rules group pick-up activity into different classifications of PMAR operation depending on the pick-up duration and the number of corresponding PMAR operations. 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. For example, the protection zone of a specific PMAR is always affected by frequent ST activities, and this indicates the high frequency of transient faults may be resulting from the existence of a SPF on the circuit.

From this prioritised mapping of potential SPF activity across the network, experts can further analyse the data using the visualisation tool for a deeper study of the pick-up data. This allows them to discern the details of PMAR operation against phases affected, evidence of earth faults and more. The process of invoking relevant rules is implemented within the KBS and fed into the visualisation tools for further analysis. The following subsection show trends of interest identified using the visualisations.

```
rule      #1: "Abnormal pick-up activity without leading to PMAR operation"
   when
                The pick-up activity's duration time is larger than 30 milliseconds
                And
                disappear without PMAR's operation
   then
                The classification is determined as FP
end

rule      #2: "Abnormal pick-up activity leading to a single trip"
   when
                The pick-up activity's duration time is larger than 30 milliseconds
                And
                cleared by a single PMAR's operation
   then
                The classification is determined as ST
End

rule      #3: "Abnormal pick-up activity leading to multiple trips"
   when
                The pick-up activity's duration time is larger than 30 milliseconds
                And
                cleared by multiple consecutive PMAR's operations
                And
                without leading to a lockout
   then
                The classification is determined as MT
end

rule      #4: "Abnormal pick-up activity leading to a lockout"
   when
                The pick-up activity's duration time is larger than 30 milliseconds
                And
                isolated by a PMAR lockout
   then
                The classification is determined as L
End
```

Figure 5.13: Rule for classification of PMAR operation

5.4.2 Evaluation of Behavioural Trends for SPF Detection

Following the classification of PMAR operation, the fault pick-up activities are divided into four groups based on their performance on the circuits where the representative trends are analysed, providing 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. Such as, when a high number of ST activities are detected after classifying the PMAR operations, the evaluation of frequency distribution could indicate the faulty conditions on account of a SPF. If those ST activities affect the same phase on the line with an increasing trend, this suggests the existence of a potential SPF.

In this research work, four statistical trends are defined and taken to evaluate the circuit behaviours for SPF detection, deduced after interviewing engineers and capturing their expert knowledge and experience. These four statistical trends respectively focus on assessing the frequency of grouped pick-up activities and the time stamp of individual pick-up activity in the group. They are [WMS+17]:

- Cumulative Frequency Distribution (CFD): displayed as an increasing line which sums the distributed frequency of classified pick-up activities to demonstrate the trend of growth. If the increasing rates keep roughly the same values for each distribution period, that means the circuit is under the acceptable conditions with natural numbers of faults. However, if the rate has an increasing value, this indicates an outbreak of pick-up activities in a particular period due to some specific reason.

- Frequency Distribution (FD): corresponding with the CFD, this details the frequency of classified pick-up activities in each period, and clearly shows the high number of activities of the particular periods related to the increasing value of rate in the CFD.

- Duration Time (DT): describes the time interval between the event registration and the end of each associated pick-up activity. This reflects the duration of a fault activity in the classified PMAR operations.

- Interval Time (IT): describes the time interval between consecutive pick-up activities. When the IT becomes shorter, that means the classified PMAR operations affecting the circuit occur more frequently. This provides the trend of occurrence for each pick-up rather than the frequency trend in the CFD.

Figure 5.14 shows an example of the trends of the four defined statistical features with considering the FP PMAR operation after the classification in the first stage (as detailed in Figure 5.13). 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.

With respect to Figure 5.14, 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 CFD (as shown in Figure 5.14(a)), this suggests the FP operations resulted from faults becoming more frequent and the corresponding FD (as shown in Figure 5.14(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 DT of each associated pick-up activity and determines the IT between consecutive pick-ups through calculation. Trends showing an increase in DT and a decrease in IT (as illustrated in Figure 5.14(c) and (d)) suggest the condition the circuit is worsening. This may also be indicative of the stage of maturity of the semi-permanent fault which results in an impending 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.

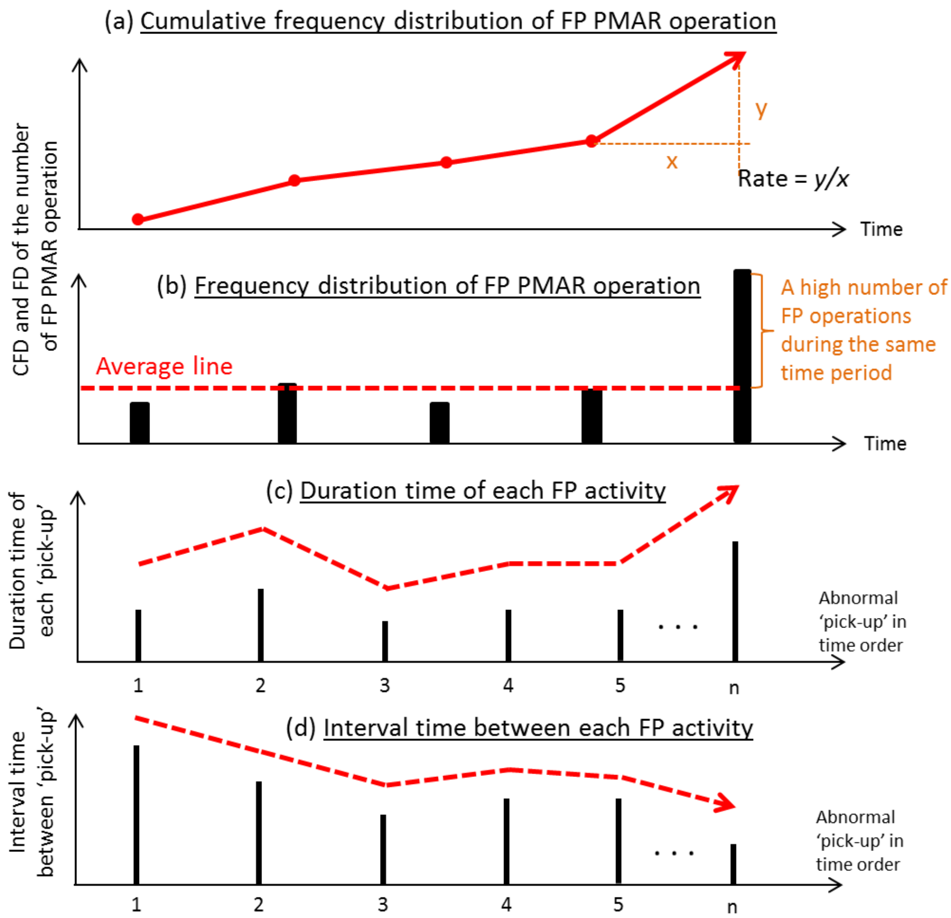


Figure 5.14: Trends of statistical features

5.5 The DSS Prototype

To efficiently interact with the DSS for automatic fault diagnosis and prognosis, and visualise the information and result of data analysis, a graphical user interface was designed to provide a staged diagnosis to the experts, and the engineers. The visualisation tool is one part of the DSS prototype to support decision making and assist users to observe information 'of interest'. The DSS prototype is included in the case studies (both fault diagnosis and prognosis) in the following sections to display the automatic fault diagnosis and prognosis process. The current version of DSS prototype has the following main components and features:

- A file importer which is capable of automatically importing all of the relevant network data (i.e. PMAR log file, PSALERTS data, PROSPER data, and NOJA list) into the MySQL database integrated with the system. The manipulated data will be stored or updated based on previous records and current analytic requirement.
- An analysis tool contains the designed fault diagnostic and prognostic functionalities automatically processing and analysing the data and returning the corresponding results to the users. The current version of the tool allows the off-line fault analysis after importing the related network data.
- A visualisation tool is capable of providing fault information of circuit behaviour with automatic data pre-processing. The current version of the visualisation tool extracts the trips and anomalous pick-up activities of a particular PMAR for a summarised or detailed view, and developed functions included in the tool are based on the requirements discussed with network experts.

Figure 5.15 shows the main user interface of the DSS, which divides the described functionalities into three sections. At the top of GUI, it is the primary section of analysis tool. On the left-hand side are functionalities relating to fault diagnosis and on the right for fault prognosis, which will be introduced in the next chapter. For the area of fault diagnosis, the first button is to automate the manual process to check the frequent short-term supply interruptions with diagnosing the potential causes recorded by maintenance staff and identify the affected PMARs. The function of the middle two buttons is to detect the SPF by classifying PMAR operations and evaluating the behavioural trends individually. The last button aims to diagnose the PMAR device faults. At the bottom of the panel is the file importer to update the relevant databases with independent functional buttons.

For the section of visualisation tool, the trip and pick-up information will be automatically pre-processed by the DSS and provided for observation based on the historical data stored in the database. That is to say, the users can view information of different PMAR at the same time depending on their interests. For example, the users can search for a particular event by viewing its time stamp, affected phases, or the time duration and current amplitude of related pick-up activity. If the PMAR data does not exist in the container or requires an update, the user should import the appropriate data for visualisation. To display the circuit behaviour from different representations, the trip and pick-up information can be viewed in various forms, such as the trips can be visualised using either a concise summary, a full distribution, or the pick-up activities that can be subsequently categorised with particular emphasis on the affected phases. Figure 5.16 demonstrates the application of the visualisation tool using data from a specific PMAR (the ID of PMAR has been hidden with consideration of network operator's privacy. This situation was applied to the case studies of fault diagnosis as well) over a time period of four years, which illustrates the trips,concentrating on fault types and invoked phases respectively in the pie charts.

Such graphical functions, although not critical to the research work or the core functionalities of the fault diagnosis and prognosis, provide an intuitive method to assist users in perceiving the fault severity and analysing the circuit condition. As already mentioned in the analysis tool, the fault diagnosis and prognosis functions are the main elements responsible for the automatic decision support, and they are demonstrated using examples in the next sections.

Figure 5.15: The main GUI of DSS with fault diagnosis and prognosis

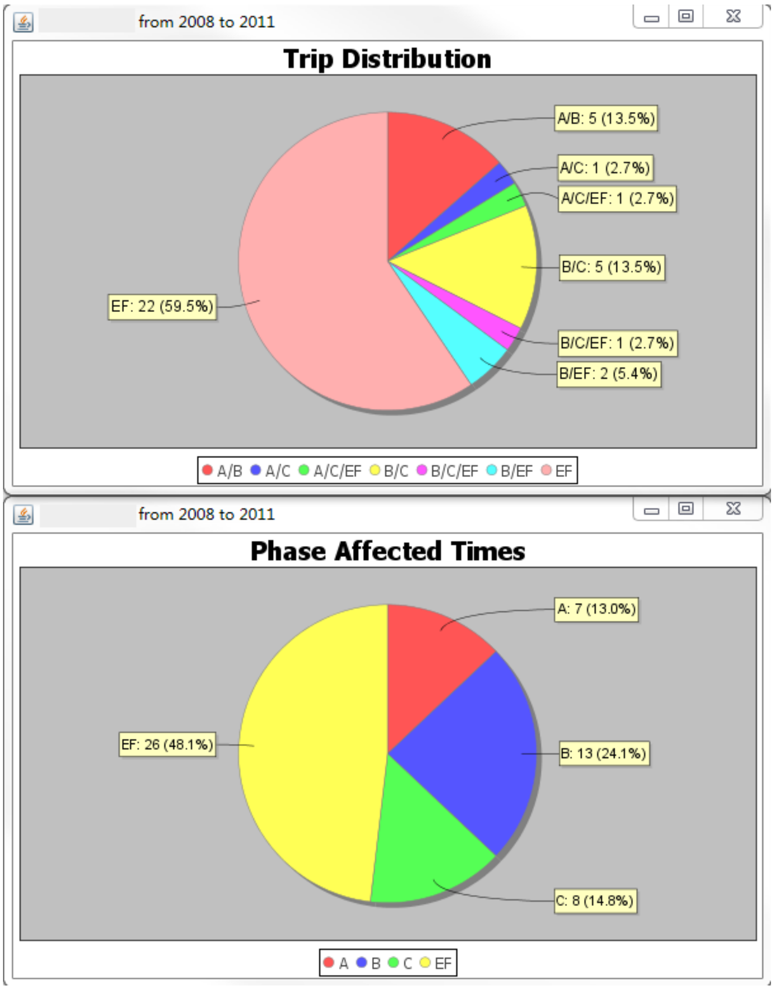


Figure 5.16: The visualisation tool to view the trip summary of a specific PMAR

5.6 Case Studies

To effectively demonstrate the working of developed fault diagnosis functionalities within the DSS, the appropriate PMARs affected by relevant faults should be chosen in advance. This case study is based around PMARs which exhibited frequent supply interruptions, detected by checking the alarm message of unsolicited openings from the morning report and identifying the previous UO records in the PSALERTS database. These frequent undiagnosed events could be the result of PMAR components and the degradation of overhead lines. The related PMAR log files were then imported into the system for further automatic analysis through the file importer. The following case studies demonstrate the automatic fault diagnosis.

5.6.1 Case Study Part 1: PMAR Device Fault Diagnosis

To test the diagnostic rules, 12 original PMAR log files were used. After importing the PMAR log files into the DSS, the knowledge-based system will automatically identify PMAR device faults and generate a report through the DSS user interface by clicking the corresponding function button. The diagnostic report contains the identified fault with a short explanation of the fault cause. For the 12 PMAR log files, all the rules described in section 5.3 have been invoked. Some PMARs were even identified with multiple different PMAR device faults, as various fault conditions matched the diagnostic rules independently. The log data to invoke the rules and the diagnostic results are shown in Appendix A. The remainder of this subsection demonstrates one example of PMAR device fault diagnosis, For this particular PMAR, the report indicates: ***Microswitch fault with the detection of an unexpected accelerated contact time***, the dialogue box is shown in Figure 5.17.

In order to validate this result of automated diagnosis, the original PMAR log file is analysed manually. From the data analysis, an unexpected accelerated contact time is calculated based on the time interval between the action signal and status change being less than the initial setting. Figure 5.10 shows the

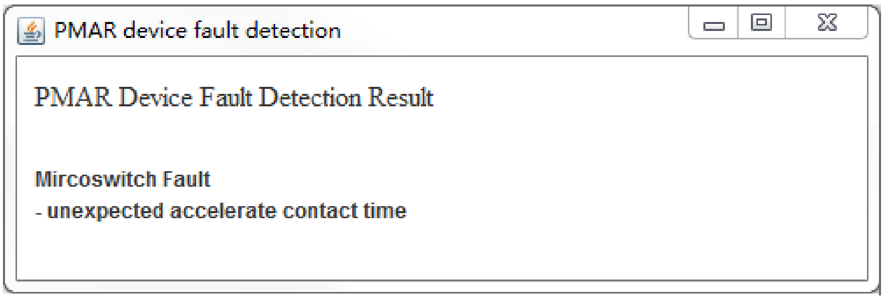


Figure 5.17: Identification of microswitch fault

01/09/2010 04:01:23:585		Trip				OC1+	B
01/09/2010 04:01:23:585		Trip	16ms			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	Uabc>	
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	29ms			AR OCEF	
01/09/2010 04:01:33:672		Closed				Driver	
01/09/2010 04:01:33:703		Pickup			End	LSD	

Figure 5.18: Validation for microswitch fault diagnosis

diagnostic rule and Figure 5.18 shows the data that activates it. In the figure, the time durations from signal ‘Trip’/‘Close’ to status ‘Open’/‘Closed’ are 16ms and 29ms respectively, both of them are less than the initial settings and result in the diagnosis. This microswitch fault could lead to additional PMAR operations, contributing to the number of short-term outages. This manual validation has been undertaken for all of the case studies in the Appendix.

This case study has shown that PMAR device faults can be identified based on the rules translated from the setting policies and experts knowledge.

5.6.2 Case Study Part 2: Semi-Permanent Fault Detection

By utilising PMAR device fault diagnosis to target the potential issues related to the frequent short-term outages detected from the morning report and historical PSALERTS database, it is usually impracticable to identify underlying anomalous circuit activity and explore the essence of problematic events, due to the volumes of data and limited information provided. This is not just because some

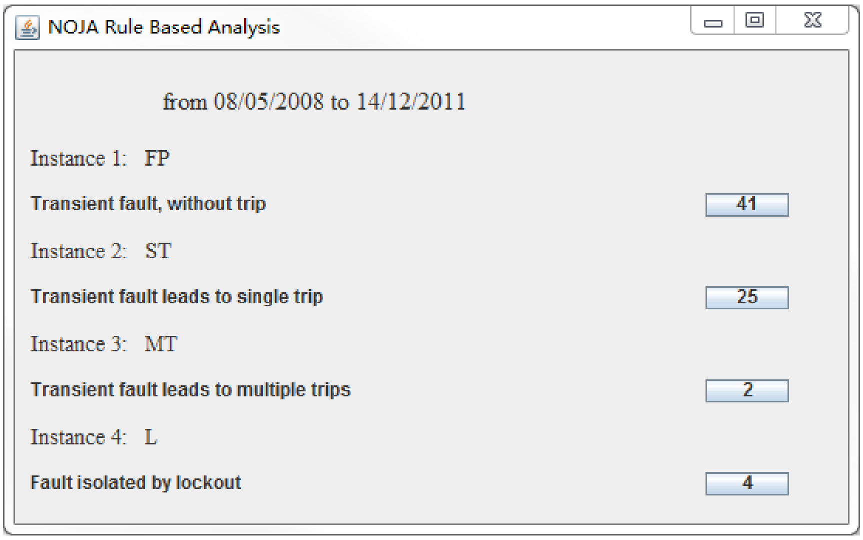


Figure 5.19: Classification of PMAR operation

types of device faults do not have direct relationships with failure of PMAR operation or circuit supply interruptions (e.g. the MPM fault with a software issue), more importantly, the PMAR device forms a singular part of the distribution network and its fault cannot represent the whole circuit behaviour. Therefore, to diagnose the nature of a circuit's fault activity, as well as the KBS automatically identifying the microswitch fault, the analysis tool within the DSS provides the control engineers with detailed information on semi-permanent fault detection through automatic data analysis and visualisation functionalities.

To consistent with the previously described analysis stages of SPF detection, Figure 5.19 illustrates the classification of PMAR operation (i.e. first step of detecting SPFs) after automatic processing with the imported specific PMAR log file. By invoking the classification rules in Figure 5.13, there are 41 **FP** activities and 25 **ST** activities, with only two **MT** activities and four **L** activities. That means in the protection zone of the specific PMAR, almost all fault activities are identified as transient faults which disappeared themselves or cleared by PMAR operations. These transient faults therefore contributed the frequent short-term supply interruptions which could have originated from the existence of SPFs.

In order to view the detail of coordinated transient fault activities and prepare for the further analysis of the trend at the second stage of SPF detection,

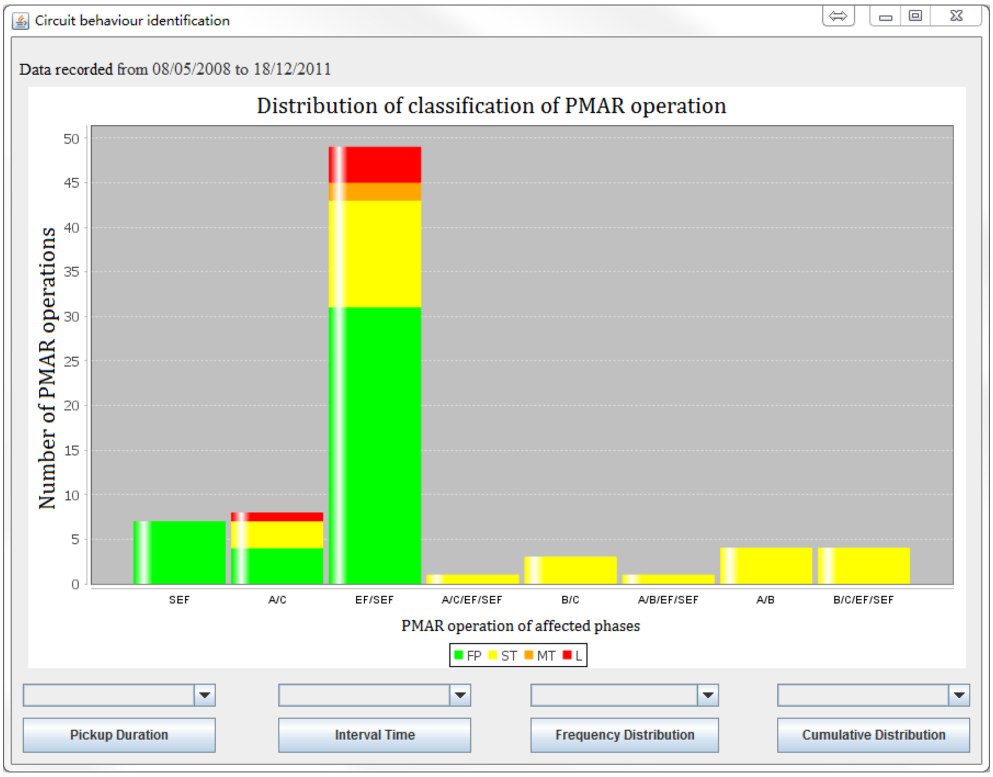


Figure 5.20: Distribution of PMAR operation

Figure 5.20 shows the distribution of these classified PMAR operations based on the affected phases of each independent fault activity. Concerning the figure, the different fault phases are listed with numbers and classifications of PMAR operation, marked by distinct colours (i.e. green, yellow, amber and red respectively represents the classification of FP, ST, MT and L activities). 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, which covers all of the fault scenarios (i.e. the different classification of PMAR operation). Among the EF/SEF fault events, there are 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. The potential reason of the EF/SEF could be flashovers occur on the polluted insulator, this leads to the insulation layer is punctured, and the weak point grounds earth through the arc. Therefore, the flashover of insulators will result in semi-permanent faults.

To confirm the existence of SPFs, the behavioural trends of the EF/SEF activities can be evaluated to assist the decision makings. At the bottom of the

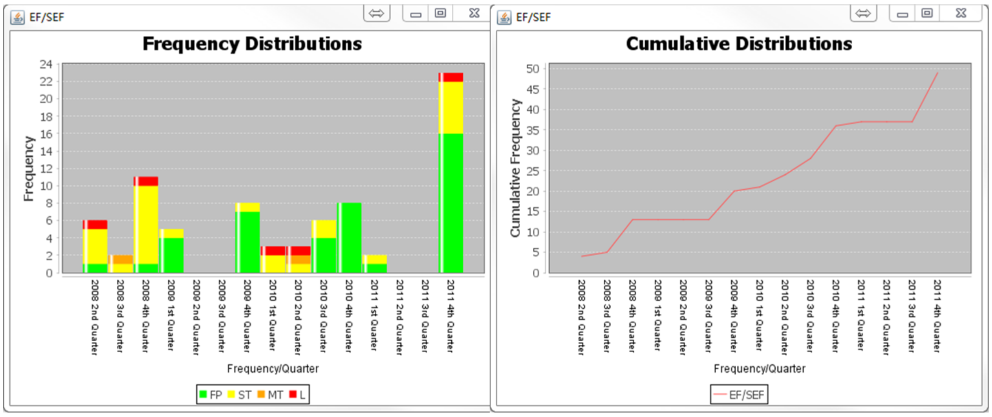


Figure 5.21: Distribution of PMAR operation

frame in Figure 5.20, the control engineers can expand the details of the key features (i.e. CFD, FD, DT, IT, described previously) describing the selected fault type. Figure 5.21 gives the detailed information of EF/SEF's FD and CFD.

With the frequency distribution shown in Figure 5.21, the EF/SEF fault always disappear without any PMAR operation or is cleared by a single attempted reclosure (represented by the green and yellow colours) resulting in the number of FP activities increasing during the same time period. Though the FP fault scenario does not lead to any PMAR operations, the growing number (especially the last distributed period (i.e. 4th quarter in 2011) in the demonstrated window) 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.

In this part, the presented case study is an example of detecting the SPF based on the appearance of frequent transient EF/SEF faults either resulting in short-term supply interruptions or otherwise. This highlights how the DSS processes the detection with the implemented functionalities of automatic classification and behavioural trend evaluation. To assist the engineers to make decisions on the analysis results after automatically invoking diagnostic rules, the integrated visualisation function, as part of the visualisation tool enhanced the confirmation of the existence of SPFs.

5.7 Conclusion

Automatic diagnosis of PMAR device faults and detection of semi-permanent faults is a challenging task, and uses the unsolicited opening records in the morning reports and available relevant PMAR log files. In this chapter, it has been demonstrated how such challenges can be overcome by the decision support system that has been developed in this research using a knowledge-based technique. The knowledge that has been utilised for constructing the rule base of the KBS has been also introduced. In this chapter, the two main fault diagnosis functionalities are implemented by deploying the KBS through generating the appropriate rule-bases, based on the knowledge translation from experts' knowledge and experience, in addition to protection and device setting policies.

The functionalities and the associated benefits offered by the DSS have been demonstrated through two illustrative and realistic case studies (i.e. PMAR device fault diagnosis and semi-permanent fault detection), based on the developed visualisation tools, as part of the design of the system. Although the two case studies do not demonstrate the entire functions of the DSS within its prototype, they allow the review of the whole process of the described fault diagnosis functionalities, providing positive results after automated analysis. This proves the designed KB DSS with the implemented rule-bases is effective in diagnosing the underlying conditions of PMAR devices and circuits, offering a satisfactory and flexible solution for the decision support tasks.

Chapter 6

Automatic Fault Prognosis Functionality with Case Studies

6.1 Introduction

With respect to the designed DSS that assist engineers in recognising PMAR condition and circuit behaviour, Chapter 5 described the fault diagnostic functionalities of identifying the PMAR device faults and detecting potential semi-permanent faults present on the overhead lines. These were responsible for the circuit's unsolicited openings. In addition to this functionality, this chapter details the research conducted into developing a prognostic capability which can predict potential PMAR operations (or evolution of SPFs into permanent faults). It grades the circuits' 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 which are required to repair damage from permanent faults.

This chapter describes the processes of how to use the PMAR data to generate the predictive rules to provide the 'early-warning' of PMAR operations, following the general path of data science technologies (e.g. data mining, clustering, and denoising techniques, etc.) to analyse data. The data analysis sections include the PMAR data's preparation, segmentation, visualisation, interpretation and testing. The proposed structured method of data analysis can be reapplied to other distribution automation and operational devices. In developing this prognostic functionality, the DSS implements the prognostic rule base into the KBS as well. Finally, the section will demonstrate the case study associated with the implemented fault prognostic functionality.

6.2 Data Selection and Preparation

The objective is to create predictive rules which can provide 'early-warning' of future PMAR operations. Hence, the first step of the work is to identify a specific fault scenario or a particular anomalous condition on the overhead line which has the capability of predicting potential PMAR operations in the future. In order to target the specific fault scenario or condition with the predictive capability, the categories of circuit behaviour should be assessed and classified in detail. In Chapter 5, the research work has already identified four classifications. Three of these (i.e. single trip (ST), multiple trip (MT), and lockout (L)) mean that the PMAR has caused a trip operation. The Fault Pickup (FP) classification means the PMAR has not caused a trip. Trip situations are indicated via ST, MT and L. Therefore, the challenges is to find enough data that allows the prediction of:

- ST, MT or L from FP events;
- ST or MT events from other ST/MT events;
- L from ST or MT events;
- Looking in the other direction, ST or MT events from L events; or
- L from other L events.

Through applying the enumeration methodology (when making inductive reasoning, if individually examined all possible cases of certain types of events, which draw general conclusion. This conclusion is reliable, this inductive method is called enumeration method) [Tho03], Table 6.1 lists all the possibilities of predicting PMAR operations based on different classifications, where the predictor

Possible patterns	Predictor	Predictand
1	FP	ST/MT
2	FP	L
3	ST/MT	ST/MT
4	ST/MT	L
5	L	ST/MT
6	L	L

Table 6.1: Possible fault patterns with predictive capability

means the potential classification of PMAR operation with predictive capability, and the predictand represents the corresponding predicted values.

In the table, the classifications of ST and MT are combined as one category of fault scenarios, because they are essentially different from L. The L event permanently opens the auto-recloser to isolate the fault, but the ST and the MT reclose after automatically clearing the fault following different numbers of reclosure attempts. As presented, six listed patterns state the potential prediction. For example, Pattern 1 describes the FP could be capable of predicting ST/MT. Therefore, in order to identify the specific fault scenario (classification of PMAR operation with predictive capability), the relationship between a predictor and the corresponding predictand of each pattern will be evaluated. If a predictive relationship can be detected from one of the listed patterns, then, the information of this particular predictor can be extracted and refined to generate the predictive rules for fault prognosis.

For identifying the possible predictive relationship, in this research work, 12 PMAR log files have been taken into consideration. These data files represent the pick-up activities captured across seven circuits, spanning a period of up to five years. In the analysis, the pick-up activities were first classified into four different levels of PMAR operations (i.e. FP, ST, MT and L) based on invoking the rules in Figure 5.13. Then, all the classified activities were used to match with the six listed fault patterns by detecting the preceding classified PMAR operation of each activity, this process was operated manually through the developed visualisation

Matched pattern	Matched amount
FP to ST/MT	102
FP to L	4
ST/MT to ST/MT	18
ST/MT to L	11
L to ST/MT	0
L to L	0

Table 6.2: The matched patterns

tool integrated within the DSS. As a result, a total of 211 FP activities, 120 ST/MT activities and 15 L activities were derived from the analysed data set. Table 6.2 shows the number of each pattern after matching process.

From the view of the table, the classified FP activities precede most of the ST/MT activities (102 of the total 120 activities), which followed by the other 18 ST/MT activities and 11 L activities. Meanwhile, the FP activities directly precede the rest of (4) L activities, so there should be no matched pattern which demonstrate the L activity preceding the ST/MT or L. This can also be verified from the table list, there is no matched Pattern 5 (L to ST/MT) or Pattern 6 (L to L). It is apparent that, the high number of PMAR operations (includes ST, MT and L) are preceded by the FP activities (defined as a pick-up duration of greater than 30ms which has not yet led to a PMAR operation). This indicates the potential underlying relationship between the FP activities and PMAR operations, which may provide information to produce the rules with the capability of predicting future PMAR operations. Therefore, the work then focuses on analysing the FP activities to derive the prognostic rules with the following detailed data science technologies.

6.2.1 Feature Preparation of FP Activity

With the visualisation of 211 FP activities and their following PMAR operations (i.e. ST/MT, L), an 'interesting' phenomenon was detected. Usually before the appearance of PMAR operations, the preceding FP activities would occur multiple times in the period of from one week up to two months. Hence, before analysing the data of FP activity and extracting the features for generating predictive rules, the 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), which may be the explanation of this particular phenomenon.

After the pre-processing of grouping the FP activities, a total number of 100 FP groups were derived from this training data set (with 211 FP activities). For identifying and summarising the predictive information from the grouped FP activities, the features for describing the FP activity should be defined and categorised. Based on the characteristics of the FP activity and the limited FP information extracted from the original PMAR log files, this work defined five features to describe a FP, and those were used for the predictive algorithm to analyse [WMS+17]:

1. Number of Total Pick-ups (NTP), which is counted 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 the time interval between two consecutive fault pick-ups in the group of FP activity (i.e. True for the decreasing of the time interval, False means the trend of the time interval is not decreasing).

4. Average Current Amplitude (ACA), which is the average current amplitude of FP activities in each group.

5. Average Pick-up Duration (APD), which is the average pick-up duration of FP activities in each group.

With these defined five features, a FP group can be clearly represented. In particular, the feature of TTT reveals the relationship with the following PMAR operation as well, which may be one of the predictive factors when producing the rules for predicting future PMAR operations. Through the computation, in which 100 FP groups are represented with five listed features, the parameters are prepared into a data sheet and subjected to the data mining and clustering process discussed in the following sections.

6.3 Data Segmentation and Visualisation

For the purpose of utilising the data mining techniques to extract predictive capability from the defined features, then the predictable features should be identified to support the generation of prognostic rules. Since the PMARs 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.

6.3.1 The Selection of Appropriate Methodologies

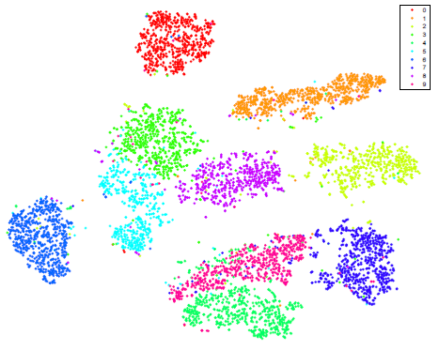
As described in the introduction of AI technologies in Chapter 3, the K-Means algorithm is a clustering technique used for data mining which is simple to implement and straightforward to interpret. It partitions similar unlabelled data (i.e. data without defined categories or groups) into a pre-set number (K) of clusters [Alp10]. When compared with alternatives, such as the hierarchical clustering algorithm [MC11], the performance of K-Means is better if the dataset is small (i.e. less number of clusters), which presents the K-Means could provide increased accuracy than the other algorithms. 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. Although some alternatives may offer a higher quality of clustering than the K-Means, such as the self-organisation map algorithm [Abb08], the algorithm is much harder to implement efficiently and has much more parameters to set. Also, the clustered results are more complex to interpret for use. However, the K-Means becomes a great solution for the pre-processing in this study, with its simplest models. Therefore, the application of a

K-Means algorithm enables segmentation of the data into distinct clusters where the specific clusters could be considered indicative of distinct PMAR operating conditions.

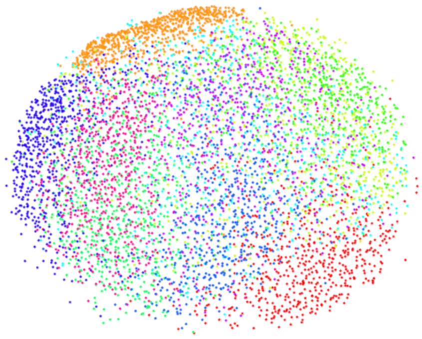
Due to the five features that are taken into the clustering process, the clustered results would be displayed into five dimensions, where each vector represents the corresponding feature. To visualise the clustering output from the K-Means algorithm adequately, a dimensionality reduction technique is required to process the clustered data. Dimensionality reduction offers a meaningful representation by transforming high-dimensional data with reduced dimensionality. Ideally, the reduced dimensionality should represent the essence of the data (i.e. corresponds to the intrinsic dimensionality of the data), where the compressed representation has the minimum number of the parameters that accounting for the observed properties of the data.

Traditionally, dimensionality reduction was performed effectively with real-world data by the linear techniques, such as Principal Components Analysis (PCA) or classical scaling methodology [VDMPVdH09]. However, the present real-world data is more likely to form a highly non-linear manifold. Due to previous linear techniques cannot efficiently manage complex non-linear data, consequently, in the last decade, a significant number of non-linear techniques have been proposed to reduce dimensionality, such as Isomap, Locally Linear Embedding (LLE), Sammon mapping, and t-Distributed Stochastic Neighbour Embedding (t-SNE) [VDMPVdH09].

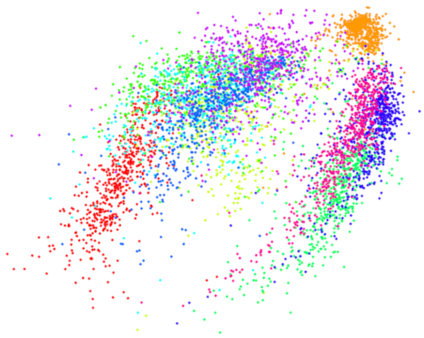
In this research, the five defined features contain non-linear data (e.g. the data of ITT). Therefore, the non-linear technique for dimensionality reduction will be adopted in this study. The t-SNE transforms high-dimensional data into low-dimensionality (two or three dimensions) [MH08], based on two main processing steps: firstly, t-SNE constructs a high-dimensional probability distribution between each object (i.e. data), which makes similar objects have higher probabilities to be selected, while dissimilar objects have a lower probability of being selected. Secondly, t-SNE constructs a low-dimensional probability distribution of these objects as well, Meanwhile, compared to other non-linear techniques, the



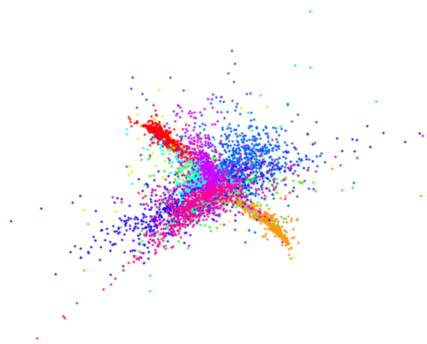
(a) Visualization by t-SNE.



(b) Visualization by Sammon mapping.



(c) Visualization by Isomap.



(d) Visualization by LLE.

Figure 6.1: Visualisation of 6000 handwritten digits with different non-linear clustering algorithms [MH08]

t-SNE not only can visualise the clustering map as a scatter-plot more clearly, but it can keep a higher accuracy when dealing with small data sets. Both of the benefits are suitable for speciality of this research work. For instance, Figure 6.1 shows the comparison of the results of the experiments with t-SNE, Isomap, LLE and Sammon mapping on one publicly available dataset, which contains 6000 grayscale images of handwritten digits (from 0 to 9).

In Figure 6.1, 10 different colours represent 10 classes. Here, the class information is only used to select a colour for the map points, not to determine the clustering coordinates of the map points. Therefore, the colouring is a way to assess how well the map preserves the similar points (i.e. the same digits) within each class. Obviously, the results reveal the perfect performance of t-SNE compared to the other techniques. Sammon mapping clusters the point into a construction like a 'ball', in which only circumjacent classes are clear relatively.

Isomap and LLE produce poor solutions with the existence of large overlaps between the different digit classes. Moreover, with the detailed inspection of the t-SNE map, the apparent boundaries of each class reveal that the local structure of the original data is captured and presented effectively. As a result, in the research work, 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.

6.3.2 t-SNE Visualisation of K-Means Clustering

As discussed previously, for processing the K-Means clustering, the number (K) of clusters should be first preset. To visualise various clustering results to support with efficiently extracting predictable information, this study tests a range number of clusters to segment and visualise the clustering of FP instances, to prevent potential false clustering of the real similar FP instances. Figure 6.2 demonstrates the comparison of cluster distributions of feature vectors as the number of clusters in the K-Means algorithm is increased. In this figure, study shows four clustering results with pre-determined numbers of cluster (i.e. 7, 10, 15 and 20), because these are quite different from each other. 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 generating the predictive rules.

As shown in Figure 6.2, the t-SNE visualises the results in a 2D representation from compressing the five-dimensional vectors while keeping the properties of the data. Each cluster is represented by distinct colour, and 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, inside and outside the circled area. While the membership (the similarity between each data point) of the data points in the clusters outside the circle changes significantly as the number of initialising clusters increases, the

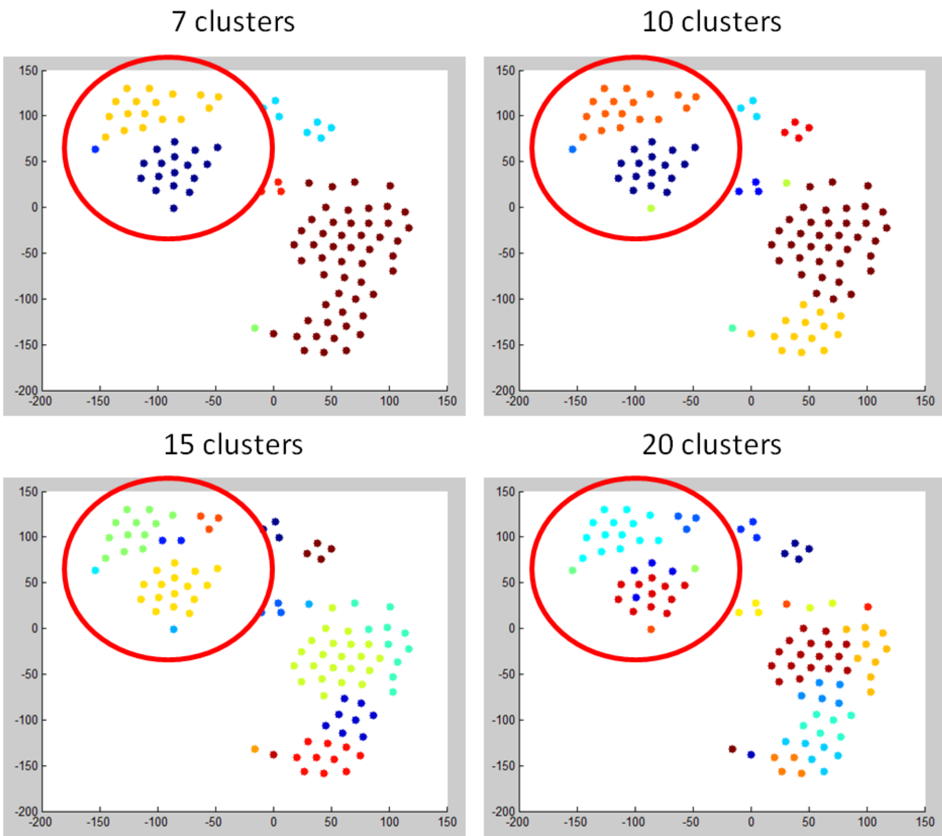


Figure 6.2: t-SNE visualisation of K-Means clustering with different numbers of clusters

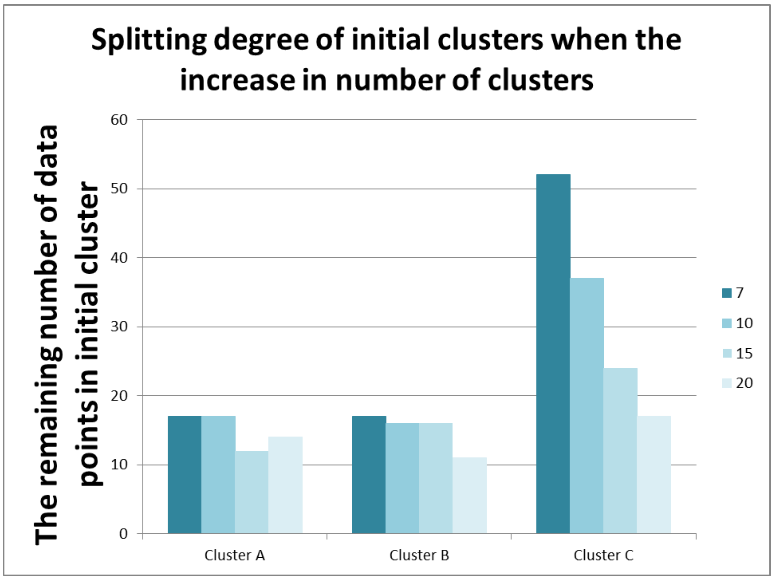


Figure 6.3: Splitting degree of initial clusters with the increase in number of clusters

membership of the data points inside the circled area does not [WMS+17]. It is evident that the distributed data in the circled area is more stable and consistent, and demonstrates a higher correlation coefficient. Instead of inspection, to validate the stability and consistency of the data in the circled area, this work will evaluate the splitting degree of data points in the clusters as the number of initialising clusters increases. With the increase in number of clusters, the clustering constraints will also increase accordingly. That means clustering conditions are more and more meticulous, which can lead to the data points (with lower similarity) being split from the original cluster and re-clustered into a new class. In other words, when the number of data in a cluster is less likely to be greatly affected (i.e. be split) by the increase in number of clusters, this indicates the data in such cluster is relatively stable and consistent. With the view of Figure 6.2, when the number of clusters was set at 7, the circled area had two apparent clusters (coloured by yellow and blue), one distinct cluster (coloured by coffee) outside the circuit. Figure 6.3 shows the splitting degree of data points in these three clusters by displaying the remaining number of data points in the initial clusters after each clustering process with increasing the number of clusters, where cluster A, B, C represents the yellow, blue, coffee cluster respectively.

Concerning the comparison of splitting degree within three clusters, the cluster

A and B are obviously more stable and consistent than cluster C, because the decreasing trend of remaining number of data points in cluster C is larger than those in others. This means that the feature vectors within the cluster A and B 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.

6.4 Data Interpretation and Testing

In order to generate rules to predict future PMAR operation, the features of NTP, ITT, ACA and APD are utilised to determine the TTT to predict a flexible time window that engineers are available to take actions. The coherent distributed data points in the circled area are analysed with deriving the values of different features corresponding to each FP group. Then, the identified range of the features' values could be used for generating the thresholds of the predictive rules.

6.4.1 Data Denoising

As the consistency indicated in the circled area, with evaluating the values of five features in the original dataset, the similarity of these data points was proofed from the data layer. To better observe how the points within the cluster differ with respect to the different features characterising them, a parallel chart shown in Figure 6.4 describing the distribution of values of the features characterising the cluster representing PMAR operation. These are plotted for the 33 data points which consistently remain members of the cluster shown as the circled area in Figure 6.2.

The chart in Figure 6.4 pplots the range of values, and each continuous line

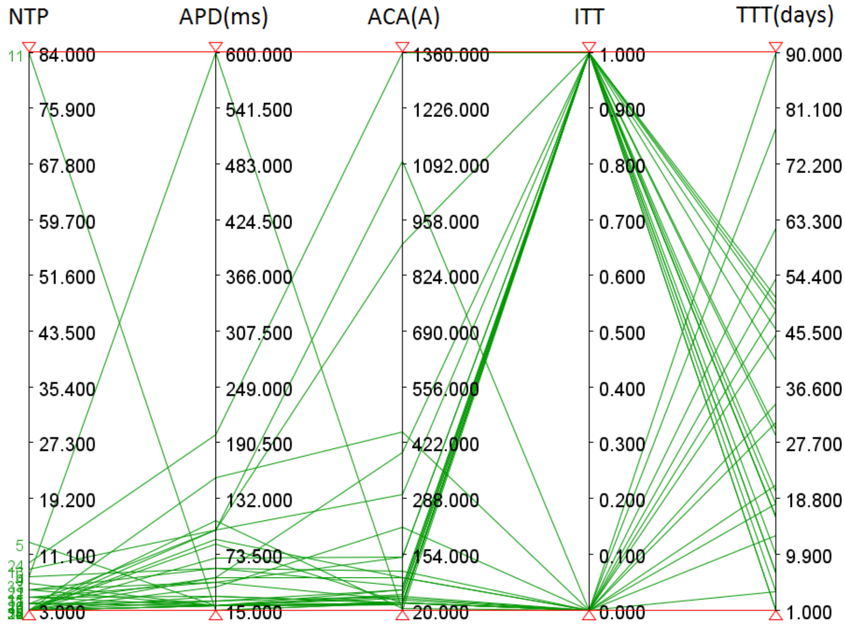


Figure 6.4: Parallel coordination plot of the segmented data

represents the value of feature in a PMAR record, where the ranges are set by the maximum and minimum values related to the features. From the chart, it is clear 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. Figure 6.4 shows that a sequence of fault activity showing these features will result in a trip will occurring 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.

To eliminate noise among the data points in the circled area, an approach of clustering the coherent data together except for the noisy data is required. Therefore, at this stage of developing the automatic fault prognostic functionality, one of the clustering algorithms should be adopted to remove noisy data. Compared to other alternatives, such as the partitioning methodologies or hierarchical methodologies, the density-based approaches have a stronger ability for noise reduction. Because the idea of these density-based methods is that when the density of the points in the region is greater than a certain threshold, all these points are attributed to a class. Data which is not in the region, and does not belong to any other regions of classes will be quickly and effectively recognised as the noisy data.

In this research work, one density-based clustering algorithm named Ordering Points To Identify the Clustering Structure (OPTICS) [ABKS99] was used to remove the outlier (noisy) data points from the encircled data. The reason for choosing the particular technique is that the OPTICS reduces the error for misidentification of noisy data, comparing with other density-based methods, such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN). These methods sometimes could treat the low-density data as noisy data. However, the OPTICS overcomes this with its function of detecting varying density clusters. Meanwhile, in the OPTICS algorithm there is an integrated function called reachability distance, which has the capability of graphical representation that highlights the outlier data besides the clusters. In this research, it can be perfectly used to detect and filter the outliers through the application of a reachability distance function.

Overall, OPTICS aims to segment data into clusters of varying density by separating the regions of high density and low density. The density means the number of points within a specified radius. In principle, the work of OPTICS is to target the core point p by using the defined density (i.e. the number of points *MinPts*, and the specified radius r). A point p is a core point if at least *MinPts* points are found within distance r. If the number of point within r is larger than *MinPts*, another radius *r'* will be defined, where the number of points is equal to *MinPts*. Finally, the points are not within *r'* but within r, the distances from core point will be calculated and called reachability distances. When the process ends, the data with higher reachability distance will be considered as noisy data. Figure 6.5 shows reachability distance function in the OPTICS methodology finds the outliers in the segmented data [ABKS99].

As illustrated in Figure 6.5, 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. For the centre of the cluster, it should have the lowest reachability distance, and a higher reachability distance represents a greater distance from the cluster centre. Hence, the outlier data must have a

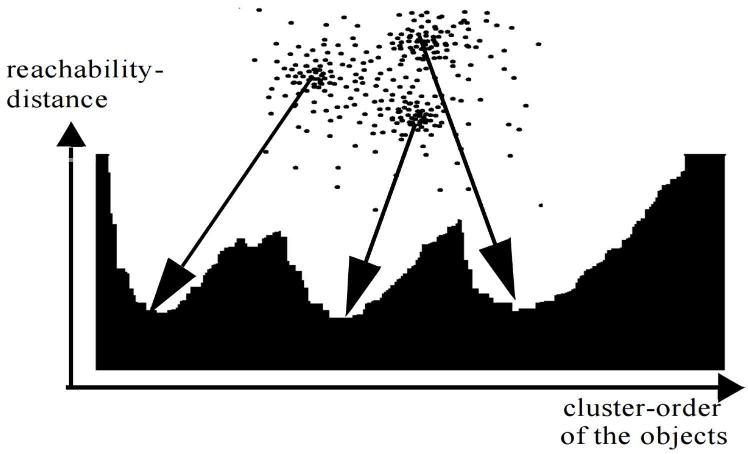


Figure 6.5: Reachability distance of the cluster-ordering [ABKS99]

high reachability distance that can be indicated obviously from the plot. Based on this, OPTICS studied the encircled data and plotted the noisy data using the presentation of the reachability distance function. Figure 6.6 demonstrates the ability of the reachability distance function to locate the outliers in the circled area via the OPTICS technique.

With respect to the 33 analysed data points in the encircled cluster, 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 Figure 6.6. These eight data points are then removed as outliers. To produce the predictive rules the other 25 data points are then subjected to the parallel chart to define the thresholds.

To validate the accuracy and efficiency of the noisy data reduction (which was achieved through utilising the reachability distance function integrated with the OPTICS, as shown in Figure 6.6), the research work displays the cumulative frequency distribution of each individual feature's value with a comparison of before and after noise reduction, as demonstrated in Figure 6.7, where the red line and blue line respectively represent the cumulative frequency distribution of the feature's value with and without noisy data. With the view of the Figure 6.7, it is clear that the distribution of each feature's value without noisy data is more concentrated than that of with the noisy data, this leads to the cumula-

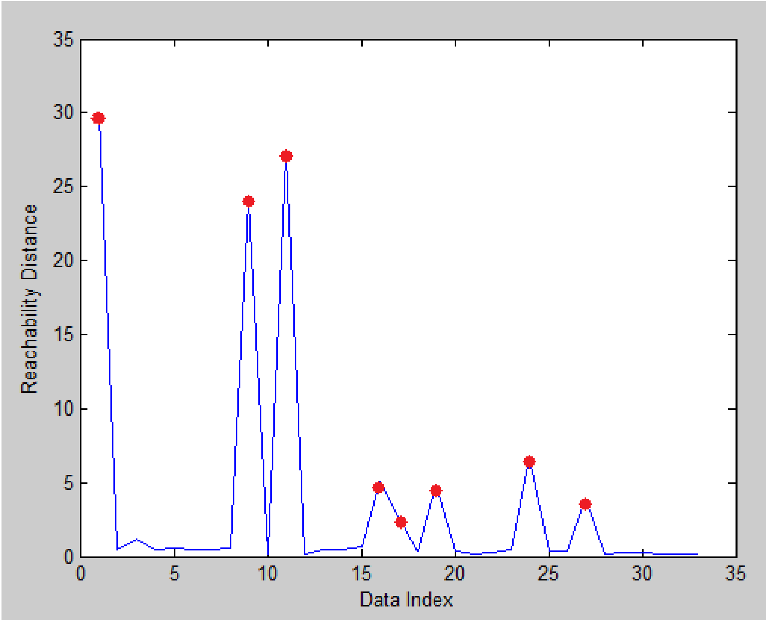


Figure 6.6: Visualisation of noisy data with OPTICS algorithm

tive frequency fast get to 1. Therefore, the reachability distance function of the OPTICS did the efficient work in the study. In this validation, the ITT feature is not included, because its value is Boolean value, which is not suitable for this distribution method.

6.4.2 Predictive Rule Implementation

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 PMARs operation. Figure 6.8 demonstrates the new parallel coordination plot of the five features after the removal of outliers.

Figure 6.8 shows the value ranges of the features NTP, APD, and ACA which form the threshold for the predictive rule shown in Figure 6.9. Furthermore, Figure 6.8 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

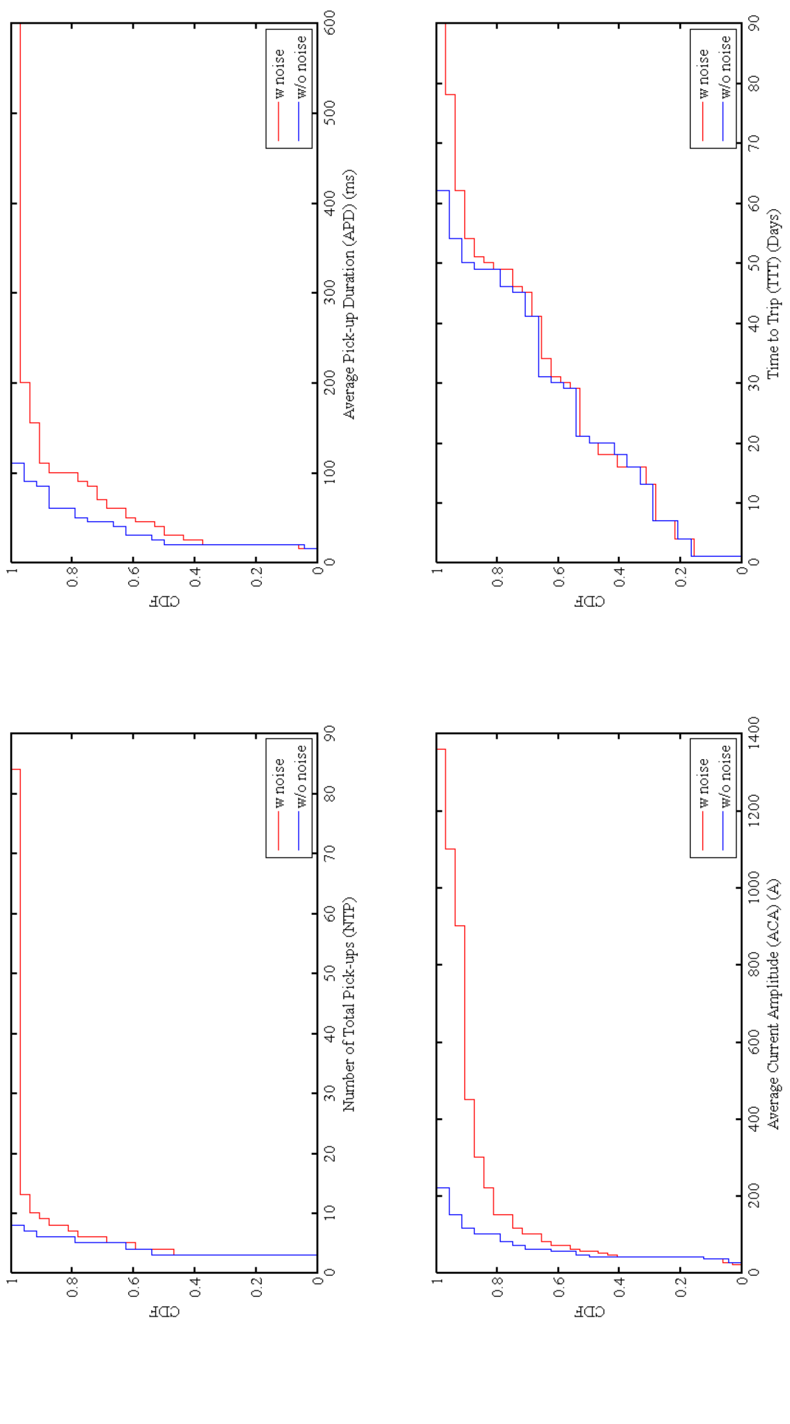


Figure 6.7: Cumulative frequency distributions of the features' values with and without noisy data

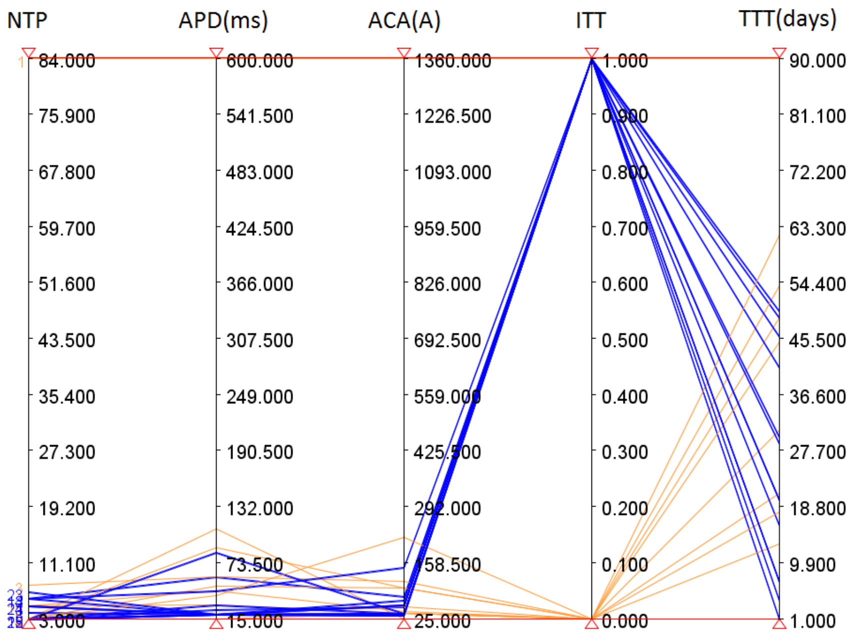


Figure 6.8: Parallel coordination plot of features without noisy data

```
rule    "General FP group for prediction"
    when
        A FP group is detected
        And
        3 < number of pick-ups < 8
        And
        15ms < average pick-up duration < 120ms
        And
        25A < average pick-up current amplitude < 220A
    then
        There will be a trip within 2 months of the last pick-up activity
end
```

Figure 6.9: The prognostic rule for predicting future PMARs operation

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 (the variety of the ITT feature's value almost will not affect the distributions of the other features' values) and consequently is not included in the derived rule in Figure 6.9. Meanwhile, this unnecessary feature explains the distinction of two main clusters (represented in two different colours) in the circled area. Compared to the previous parallel chart with outlier data, the rule is also more precise in terms of the predicted TTT, which downsizes the predictive duration to 2 months from previous 3 months.

```
rule      "Specific FP group (EF/SEF) for prediction"
    when
            A FP group only with EF/SEF pick-ups is detected
            And
            3 < number of pick-ups < 6
            And
            15ms < average pick-up duration < 100ms
            And
            25A < average pick-up current amplitude < 80A
    then
            There will be a trip within 55 days of the last pick-up activity
end
```

Figure 6.10: The prognostic rule with focusing on the detection of EF/SEF pick-ups

Through the visualisation function of pick-up information (described in previous section of DSS prototype as shown in Figure 5.15), the 25 predictable data points have been reviewed based on the level of affected phases. There are 19 FP groups with only the pick-ups of the Earth Fault and Sensitive Earth Fault (EF/SEF), and another 3 FP groups combined with anomalous activities of the EF/SEF and the phase to phase fault. Clearly, the EF/SEF is the main contributor in the category of FP activity leading to transient supply interruptions without PMAR operations. Then, the research extracted all the 19 FP groups with only the pick-ups of the EF/SEF events and aimed to generate another predictive rule to obtain more precise predictive results. Figure 6.10 displayed the prognostic rule which focuses on detecting the FP group with the EF/SEF pick-ups to predict the future PMAR operations.

As shown in the specific prognostic rule, the criteria of redefined thresholds are based on the updated maximum and minimum range values of the features. The predicted TTT is reduced by 5 days (i.e. within 55 days) as well. Both of the rules are concentrating on predicting future PMAR operations with detecting the appropriate FP groups. Due to the limitation of the actual network data available at present, the research generated one general and one specific rule for fault prognosis, the processing of generating a prognostic rule is repeatable, and the threshold in the rule would be more precise in terms of predicting future PMAR operation, as more DA data becomes available. To validate the produced rules, the rules will be tested also with actual network data, then applied and written to the designed DSS.

6.4.3 Data Testing

For validating the prognostic rules, new unanalysed original data will be fed into the designed DSS to invoke the corresponding prognostic rules. In this thesis, 27 unseen PMAR log files (an independent data set, which has not been used for previous data training and testing) containing FP activities and PMAR operations were selected for analysis. Automatic processing of the historical data, revealed that there are 32 FP groups (including 22 FP groups contributed with only multiple EF/SEF pick-up activities) of fault pick-up activities that do not lead to any trips or lockouts within 1 month which 'fire' the predictive rules in Figure 6.9 and Figure 6.10, respectively suggesting that a a trip will occur in 2 months or 55 days.

Through the 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 sensitivity or success rate of the rule is 84.4%. Compared to the specific rule focusing on EF/SEF pick-ups, although the predicted time to trip is only more precise for 5 days, 20 included in the 22 previous detected EF/SEF FP groups improved the sensitivity of prediction to 91%. This analysis of utility data verifies that the methodology adopted can be used to develop accurate predictive capabilities. Further predictive rules with more sensitivity or success rate can be created using this approach as more data is collected by the network operator.

6.5 Case Studies

To effectively demonstrate the operation of both of the prognostic rules implemented in the DSS, two different original PMAR log files are chosen to respectively provide the positive analysis. Meanwhile, in order to ensure the completeness of validation process, the updated log files of same PMARs should be extracted with related information to verify the prognostic results (i.e. the predicted time to trip). This case study is based around the PMARs which were affected by

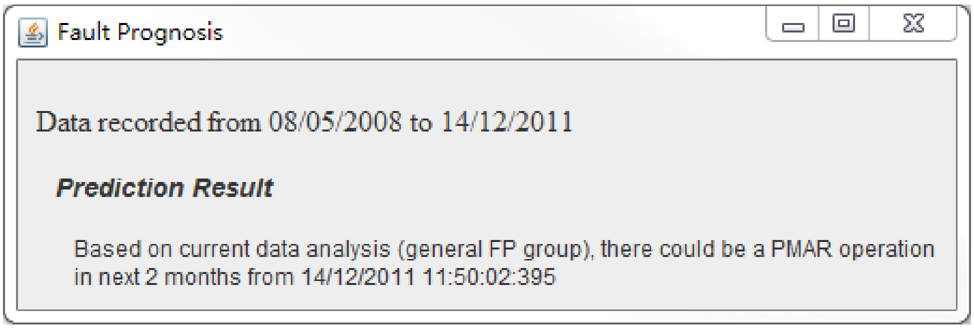


Figure 6.11: Report of fault prognosis with firing general predictive rule

frequent supply interruptions recorded in the morning reports. Similarly as with the demonstration of developed diagnostic functionalities, the relevant PMAR log files are required to import into the DSS to pre-process for the case study of automatic fault prognosis.

6.5.1 Case Study Part 1: Fault Prognosis with General FP Group Detection

After importing the PMAR log file, the DSS will automatically analyse the data for fault prognosis, engineers can also utilise the KBS applications to check the fault prediction report based on the prognostic rule implemented in the rule-base. Figure 6.11 shows that the KBS reports a potential PMAR operation will occur within 2 months from the date of 14/12/2011 by detecting a particular general FP group which invoking the general predictive rule, where the informed date (14/12/2011) is the time stamp of the last pick-up activity in the FP group. The system automatically alarms the user of this.

For validating the prediction, the new PMAR log file which contains the pick-up information after 14/12/2011 will be analysed manually. Figure 6.10 illustrates the data table from the updated log file. It can be seen that the time stamp of the next trip (which is highlighted) is 18/12/2011, which is recorded as a phase to phase (A to C) fault. 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

Figure 6.12: Validation for predicting future PMAR operation in next 2 months

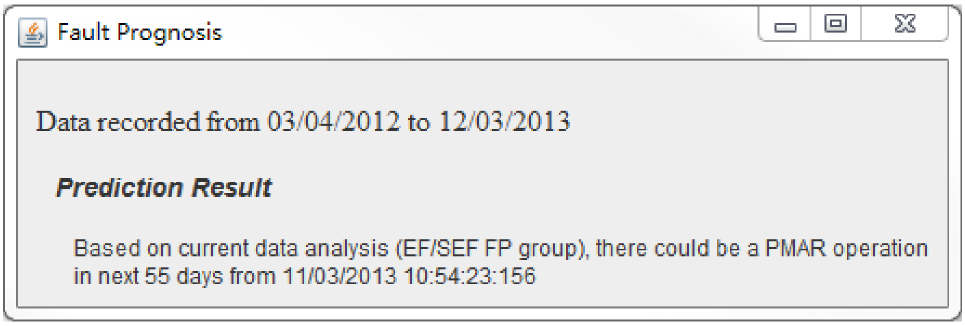


Figure 6.13: Report of fault prognosis with firing specific predictive rule

6.5.2 Case Study Part 2: Fault Prognosis with EF/SEF FP Group Detection

With respect to the demonstration of fault prognosis by detecting the general suitable FP group, in this part, another PMAR log file was analysed to predict future PMAR operations with the identification of specific EF/SEF FP group. Figure 6.13 displays the similar prognostic result with KBS report to indicate a future PMAR operation will occur within 55 days (from the time stamp of 11/03/2013). The corresponding Figure 6.14 validates the prediction result by checking the data in the new relevant log file, where a phase to phase (A to B) fault had been tripped by the auto-recloser in the next 8 days (19/03/2013), and highlighted in the data table.

In this section, the presented examples of fault prognosis exhibit the process of functionality designed and developed in the DSS, both of the predictive rules can be used to inform future PMAR operation with relevant data analysis. Based on the feasible predicted time slot before potential trip occurring within the automatic KBS report, the engineers can take actions or make decisions on dispatching the maintenance staff to the location of particular reported PMAR

19/03/2013 06:44:48:046		Pickup			Start	OC1+	A
19/03/2013 06:44:48:046		Pickup			Start	OC1+	B
19/03/2013 06:44:48:191		Pickup			Start	OC1+	C
19/03/2013 06:44:48:196		Trip				OC1+	A
19/03/2013 06:44:48:196		Trip				OC1+	B
19/03/2013 06:44:48:236		Open				Driver	
19/03/2013 06:44:48:236		AR Initiation				AR OCEF	
19/03/2013 06:44:48:236		Pickup			End	OC1+	A
19/03/2013 06:44:48:236		Pickup			End	OC1+	B
19/03/2013 06:44:48:236		Pickup			End	OC1+	C

Figure 6.14: Validation for predicting future PMAR operation in next 55 days

for future inspection and repair.

6.6 Conclusion

In this chapter, the prognostic functionality is defined and deployed within the knowledge-based system in the DSS to invoke the generated prognostic rule-base. By selecting actual PMAR log files, the case study displays the positive prediction results and shows how the DSS operates to automate the fault prognosis, and the support that can assist engineers' decision making. The source code for the rules is presented in Appendix A.

In particular, the entire process of developing the automatic fault prognostic functionality is based on the utilisation of various data science techniques to analyse the PMAR data. These are the data selection and preparation supported with using enumeration methodology, data segmentation and visualisation with the K-Means clustering algorithm and t-SNE technology, data interpretation and utilisation of the OPTICS denoising methodology. The data analysis methods can be re-applied to new PMAR devices or wider updated log files to generate more predictive rules to extend the knowledge-based system. This chapter proves the correct process of developing the fault prognostic functionalities by using the devices' data at the level of the low-voltage distribution network.

Chapter 7

Conclusions and Further Work

7.1 Conclusions

Distribution automation is deployed to ensure reliable operation and protection of power systems by reducing outages and rapidly reconnecting customers after network faults occur. Recent developments in DA systems, integrated with intelligent equipment (i.e. IEDs), have enabled automatic data analysis of logged load or fault event data. The application of data analytics and automated analysis of this data provides a picture of underlying circuit behaviour between the successive operations over a period of time, which supports post fault management and investigation. Moreover, it has the capability of informing the evolving anomalous conditions to offer feasible opportunities to take action in advance to mitigate supply interruptions (i.e. reduce CML and CI). A large volume of research has delivered corresponding diagnostic capability in related areas. For example, decision support systems utilise intelligent methods to provide fault analysis and diagnostic assistance for protection and control engineers. These practical operational systems automatically analyse the data captured from the SCADA systems or/and IEDs.

However, the complexity of the network and the vast data extracted from thousands of DA devices (and their associated settings) is an extremely challenging task. Existing relevant developed systems contain a number of shortcomings:

the systems mainly focus on fault detection and location identification by interpreting SCADA data, with support from the analysis of available IED data. For the root causes of fault events or anomalous circuit activities, there is no in-depth study which fully utilises DA devices' data. Even if research concentrated on using the IED data as the primary source to support automated data analysis to assist engineers, the proposed solutions to improve system operation and protection do not offer the prognostic capability of the potential circuit or device faults. Furthermore, no research is available on fault diagnosis by explaining the data from IEDs which are installed in the low voltage distribution networks.

In this thesis, the aforementioned problems have been addressed and the results of research into a methodology for automatic fault diagnosis and prognosis of distribution automation has been presented. It is concerned with designing a decision support system, to automatically analyse both SCADA data and PMAR data and which assists engineers in recognising underlying circuit and device behaviour. Meanwhile the system takes a further and detailed step of interpreting PMAR data to address a gap in the field of power system data analysis with the use of LV distribution network IED data. A knowledge-based system has been adopted and implemented for the proposed objectives. The KBS diagnoses PMAR faults and identifies emerging circuit and device faults (i.e. semi-permanent faults) to allow preventative measures to be taken to avoid or minimise outages. The automatic processes are based on invoking the relevant rules, which are translated from domain experts' knowledge and experience and network operation settings.

Specifically, the developed DSS contributes the design and proof of a predictive capability to identify emerging faults within a DA application. The prognostic functionality is implemented by deploying the KBS in the DSS, which develops the prognostic rules by applying the data science techniques into analysing the PMAR data. The processing steps follow the procedures of data mining including data preparation, data segmentation, and data interpretation and utilisation. Both the adopted methods of developing diagnostic and prognostic functionalities offer satisfactory solutions that have been demonstrated in the case studies.

The proposed data analysis functions have been implemented in the DSS prototype and tested using actual data from the UK distribution network. The developed prototype, as a GUI, includes the designed visualisation tool. On one hand, the visualisation tool provides the diagnostic results and supports the analysis of circuit behaviour to assist the users with decision making by observing the detailed information (fault events or anomalous activities) after automatic data processing. On the other hand, the visualisation tool searches 'interesting' fault patterns to support the generation of predictive rules to develop prognostic functionality. Meanwhile, the case studies of automatic fault diagnosis and prognosis are demonstrated through the implemented DSS prototype.

In the case studies, the proposed novel solution to timely address the use of LV distribution network data for fault diagnosis and prognosis is demonstrated by applying the actual network data against the developed rule-bases. Based on these demonstrations, it can be seen how the automatic system assists the control engineers with their fault data analysis and shows the valuable benefits of supporting the reduction of customer supply interruptions and detection of underlying asset deterioration.

During research work, it has been realised that the key challenges in developing the decision support system for automatic fault diagnosis and prognosis are the accurate selection and sufficient analysis of suitable and available PMAR data. Furthermore, the adopted methodologies of system design and data analysis to achieve the proposed functionalities can be applied to wider distribution automation devices or shifted to larger datasets, when the systems have the similar motivation of automating the analysis for fault diagnosis and prognosis. With respect to the current version of developed DSS, the utility is now in the process of initiating a project to roll this functionality out across all of their PMARs, but it can be extended over time and with greater data access.

7.2 Future Work

7.2.1 Enhancement of the KBS

In the existing developed KBS, the implemented rule-base contains the rules for three applications: rules for diagnosing PMAR device faults, rules for classifying PMAR operations and preparing visual features for detecting semi-permanent faults, and rules for predicting future PMAR operations. These existing rules for data analysis have been proven to be very useful when importing the PMAR data to execute related designed functions. However, all the above aspects can be improved for better automatic fault diagnosis and prognosis within more refined rules and can be incorporated in future.

To enhance the rule-base for fault diagnosis, the next stage of the work would investigate the definitions of further anomalous conditions of PMAR devices to cater for a wider range of tests. The new generated rules for diagnosing PMAR device faults can be achieved by interviewing domain experts and referring to relevant manufacturers' documents for validation. For full implementation of the rules to identify semi-permanent faults on circuits, the previous defined features can be extracted and amended by embedding engineers' knowledge and experience to produce new rules, so that the system can automatically detect the emerging faults and report to engineers instead of users' decision makings after manual observation.

Furthermore, the prognostic rules in the rule-base can be further extended, which could include the use of current applied data science methodologies to process more available PMAR data, in order to generate more precise rules to forecast potential PMAR operations. Otherwise, the proposed approaches could consider more features (e.g. affected phase, etc.) into the fault pick-up (FP) groups or study other classifications of PMAR operations (i.e. ST, MT, L) to generate diversified and detailed predictive rules.

7.2.2 Migration of System from Off-Line to On-Line Mode of Operation

Currently, the developed DSS can only use the data which is stored in the PMAR log file downloaded from each PMAR location as an off-line mode. The main disadvantage of such a system is that it cannot analyse the latest PMAR data to predict future PMAR operations with a feasible time period preceding the occurrence of the faults, which allows maintenance staff to take effective actions in advance.

The next step of the work can revise the system to interface with the control room which contains the on-line data sources (e.g. SCADA data or PMAR data transferred through the 'Nortech ihost' system) that can provide real-time distribution network data. Therefore, the timely data can be automatically analysed to offer the suggestions of system operation and protection.

7.2.3 Further Development of the Prototype Tool for Industrial Application

The ultimate aim of the work is to deliver an intelligent decision support system for industrial applications as business as usual. For the purpose of achieving the proposed functionalities, the DSS prototype should be further refined, developed and comprehensively tested.

The current prototype supports the fault diagnosis and prognosis with the import of the related PMAR log files of 'interest' affected area of the distribution network, which simply provides the result report to assist engineers' decision makings. Further development work is required to automatically inform the user with different monitoring levels (condition alarms with green, amber, red indications) of a wider range of detected circuits or PMARs simultaneously.

Another important aspect of improving the prototype tool is to normalise the design and compatibility, so that the tool can be easily modified or properly maintained with the future industrial requirement, such as insertion of new diagnostic or prognostic modules.

7.2.4 Comprehensive Study of Manual Process within the DSS - Roll out of the Automatic Process to the Entire System

The designed DSS analyses the PMAR data for fault diagnosis and prognosis followed by the automated manual process with investigating unsolicited openings in the morning reports. Once the development and refinement of the prototype system is complete, a comprehensive study of the manual process which focuses on analysing the SCADA alarm data (i.e. PSALERTS data) and network operator's fault diagnostic data (i.e. PROSPER data) can be performed. This could allow the successful combination with the on-line data source, and implement the entire developed DSS to a fully automatic data analysis system, from retrieving on-line relevant data to provide decision supports to the control engineers.

7.2.5 Fault Diagnosis and Prognosis Validation Based on the Analysis of Historical Data

Since the detection of semi-permanent fault and prediction of future PMAR operations are required to be verified by the evidence from real occurrences in the the network, the system cannot guarantee the fault diagnosis and prognosis with 100% accuracy. As a result, future work should also focus on fault diagnosis and prognosis validation (i.e. the confidence) based on obtaining more historical data for the analysis of the particular circuits or PMARs. While the results presented here are promising, this will assist engineers to obtain an overall confidence of for the results of the automated analysis of circuits or PMARs. For providing accurate validation function integrated with fault diagnosis and prognosis, the estimated confidence should be based on the analysis and testing of the PMAR over latest 5 years.

Appendix A

Supplementary Information for Case Studies in Chapter 5 and 6

This appendix provides supplementary details for the case studies of fault diagnosis and prognosis presented in Chapter 5 and 6. As described previously, the data used in all case studies was captured from the installed PMARs in a distribution network. The voltage level of the network is 11kV, and the PMARs generate log files containing overhead line current and voltage data sampled at 12.8 kHz. At present, over 200 PMARs are deployed in SPEN's distribution network, and each PMAR can generate tens of thousands of sets of data in the file (the size is around hundreds of kilobytes) in every month.

The Main Processor Module integrated within the PMAR records the data as a form of TXT file with encoding. To import the appropriate data into the designed DSS for automatic analysis, a corresponding decoding software named as 'TELUS' would convert the original data into the prepared data sheet in an Excel format. As a result, all of the underlying data in the case studies or used in the DSS functions is stored in an Excel format.

To ensure privacy and security of the distribution network operator data, the full original data sheet studied in the case studies will not be presented in this appendix. However, Section A.1 shows some discrete original PMAR log data to demonstrate the data format and structure of the log file, and also illustrate

Event log

Created : 04/02/2015 16:08:44

MPM #

Date and Time	Event title	Start/End	Source of	Relevant	Relevant state	Critical parameter
03/04/2012 06:46:15:807	Pickup	Start	OC1+	A		Iop, A=240
03/04/2012 06:46:15:807	Pickup	Start	OC1+	C		Iop, A=240
03/04/2012 06:46:15:952	Trip		OC1+	A		
03/04/2012 06:46:15:952	Trip		OC1+	C		
03/04/2012 06:46:15:992	Open		Driver			
03/04/2012 06:46:15:992	AR Initiation		AR OCEF		O2	Tr, s=10.00
03/04/2012 06:46:15:992	Pickup	End	OC1+	A		Max(Ia), A=1252
03/04/2012 06:46:15:992	Pickup	End	OC1+	C		Max(Ic), A=1275
03/04/2012 06:46:16:003	Pickup	Start	Uabc>			Up, kV=5.1
03/04/2012 06:46:16:023	Pickup	Start	LSD			
03/04/2012 06:46:16:023	Pickup	Start	Urst<			
03/04/2012 06:46:25:991	Close		AR OCEF		C2	
03/04/2012 06:46:26:046	Pickup	End	LSD			
03/04/2012 06:46:26:046	Toir	Start	IR			OIRM=4.00
03/04/2012 06:46:26:046	Pickup	End	Urst<			
03/04/2012 06:46:26:050	Closed		Driver			

Figure A.1: An example of fault event data

examples of different categories of PMAR data information which supports the data analysis of fault diagnosis and prognosis in the DSS (detailed in subsection 2.4.3). Section A.2 mainly provides the log data which invoked the PMAR device fault diagnostic rules, and corresponding results; Section A.3 and Section A.4 offers the source code of fault diagnostic and prognostic rules mentioned in the thesis.

A.1 Examples of PMAR Data

With respect to the description of PMAR data's structure detailed in subsection 2.4.3, this section shows the three categories (i.e. fault event, abnormal activity and device event) of data information can be extracted from the log file.

Figure A.1 displays all detailed information of a fault event registered in the log file, including the time stamp of fault, current amplitude of fault, affected phases, etc. In this special case, a Phase A to Phase C fault had been detected to trip a PMAR operation, where the gray blocks covered the ID of particular circuit and PMAR device. Figure A.2 exhibits the detailed information of abnormal activity with the detection of EF/SEF activities in a particular PMAR, but not resulting in a trip. Figure A.3 shows an example of device event in a log file with existence of alarm messages about driver module.

03/04/2012 06:52:07:821		Pickup		Start	SEF+				Iop, A=21	
03/04/2012 06:52:07:826		Pickup		Start	EF1+				Iop, A=30	
03/04/2012 06:52:07:846		Pickup		End	EF1+				Max(In), A=41	
03/04/2012 06:52:07:846		Pickup		End	SEF+				Max(In), A=41	
03/04/2012 06:52:07:896		Reset			EF1+					
03/04/2012 06:52:07:896		Reset			SEF+					
03/04/2012 06:52:17:517		Pickup		Start	SEF+				Iop, A=21	
03/04/2012 06:52:17:522		Pickup		Start	EF1+				Iop, A=30	
03/04/2012 06:52:17:532		Pickup		End	EF1+				Max(In), A=30	
03/04/2012 06:52:17:537		Pickup		End	SEF+				Max(In), A=30	
03/04/2012 06:52:17:582		Reset			EF1+					
03/04/2012 06:52:17:587		Reset			SEF+					

Figure A.2: An example of abnormal activity data

30/05/2012 13:03:36:743		Closed			Driver					
30/05/2012 13:03:36:773		Open			Driver					
30/05/2012 13:03:36:789		Pickup		Start	LSD					
30/05/2012 13:03:36:789		Pickup		Start	Uabc>				Up, kV=5.1	
30/05/2012 13:03:36:789		Pickup		Start	Urst<					
30/05/2012 13:03:36:798		Closed			Driver					
30/05/2012 13:03:36:849		Open			Driver					
30/05/2012 13:03:36:851		Pickup		Start	Uabc>				Up, kV=5.1	
30/05/2012 13:03:36:851		Pickup		Start	Urst<					
30/05/2012 13:17:06:061		Close			SCADA					
30/05/2012 13:17:06:119		Closed			Driver					
30/05/2012 13:17:06:120		Pickup		End	LSD					

Figure A.3: An example of device event data

A.2 Data and Results for Case Studies in Chapter 5

As introduced in the case study of PMAR device fault diagnosis in Chapter 5, the example in section 5.6 only demonstrates the microswitch fault diagnosis, this section provides the supplementary details of diagnostic case studies of the other four categories of PMAR device faults.

A.2.1 Data for Case Study of MPM Fault Diagnosis

After invoking the MPM fault diagnosis rule with matched data captured from the data sheet, the DSS automatically reports the diagnostic result: ***Main Processor Module fault with the detection of a number of fleeting MPM alarms***, as shown in Figure A.4. The associated data is presented in Figure A.5, where the consecutive number of "MPM fault" events exceeds 20, firing the corresponding rule in Figure 5.3.

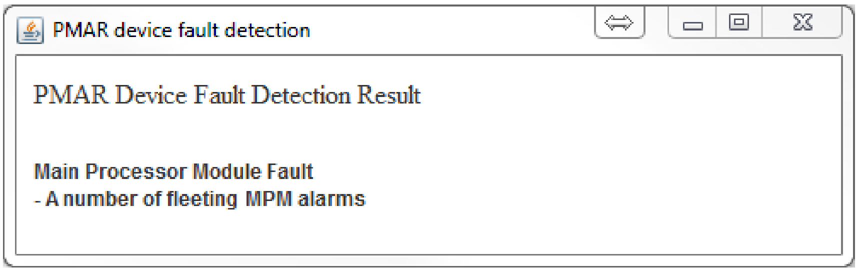


Figure A.4: Identification of MPM fault

31/07/2010 02:22:45:824		MPM fault		Start	ISC
31/07/2010 02:24:55:672		MPM fault		End	ISC
31/07/2010 04:28:10:149		MPM fault		Start	ISC
31/07/2010 04:30:19:915		MPM fault		End	ISC
01/08/2010 00:47:28:279		MPM fault		Start	ISC
01/08/2010 00:49:37:840		MPM fault		End	ISC
01/08/2010 05:34:53:476		MPM fault		Start	ISC
01/08/2010 05:37:03:161		MPM fault		End	ISC
01/08/2010 07:05:38:327		MPM fault		Start	ISC
01/08/2010 07:07:48:096		MPM fault		End	ISC
01/08/2010 20:03:46:023		MPM fault		Start	ISC
01/08/2010 20:05:55:459		MPM fault		End	ISC
01/08/2010 20:27:32:706		MPM fault		Start	ISC
01/08/2010 20:29:42:060		MPM fault		End	ISC
02/08/2010 00:57:45:216		MPM fault		Start	ISC
02/08/2010 00:59:55:238		MPM fault		End	ISC
02/08/2010 04:25:19:924		MPM fault		Start	ISC
02/08/2010 04:27:29:779		MPM fault		End	ISC
02/08/2010 05:15:03:289		MPM fault		Start	ISC
02/08/2010 05:17:13:025		MPM fault		Start	ISC
02/08/2010 05:19:22:622		MPM fault		End	ISC
02/08/2010 05:30:10:936		MPM fault		Start	ISC
02/08/2010 05:32:20:803		MPM fault		End	ISC
02/08/2010 05:47:30:751		MPM fault		Start	ISC
02/08/2010 05:49:40:554		MPM fault		End	ISC

Figure A.5: Validation for MPM fault diagnosis

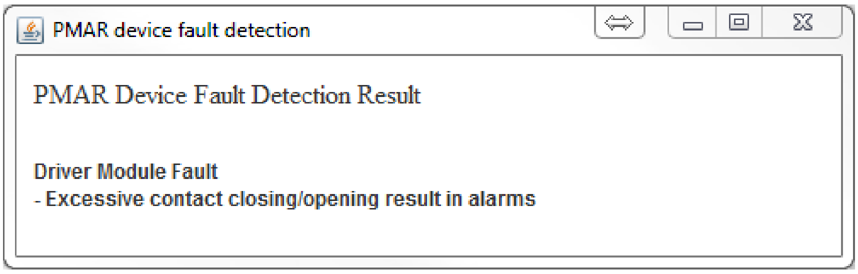


Figure A.6: Identification of driver module fault

27/08/2006 08:42:19:080		AC Off			End	UPS
27/08/2006 08:42:26:888		Trip				MMI
27/08/2006 08:42:27:105		Excessive To			Start	ISC
27/08/2006 08:42:38:477		Trip				MMI
27/08/2006 08:42:38:522		Open				Driver
27/08/2006 08:42:38:522		Excessive To			End	ISC

Figure A.7: Validation for driver module fault diagnosis

A.2.2 Data for Case Study of Driver Module Fault Diagnosis

Figure A.6 shows the DSS automatically identified a driver module fault with giving a fault diagnostic report: ***Driver Module fault with the detection of excessive contact closing/opening result in alarms***. From observing the original log data in Figure A.7, it is can be seen that the alarm message "Excessive To" fired the corresponding rule in Figure 5.5.

A.2.3 Data for Case Study of Tank Fault Diagnosis

The diagnostic report (Figure A.8)shows the message of ***NOJA Tank fault with the detection of driver status changes faster than settings***, which indicates the DSS identified a PMAR tank fault. Figure A.9 illustrates the associated data met the conditions of the tank diagnosis rule, described in Figure 5.7.

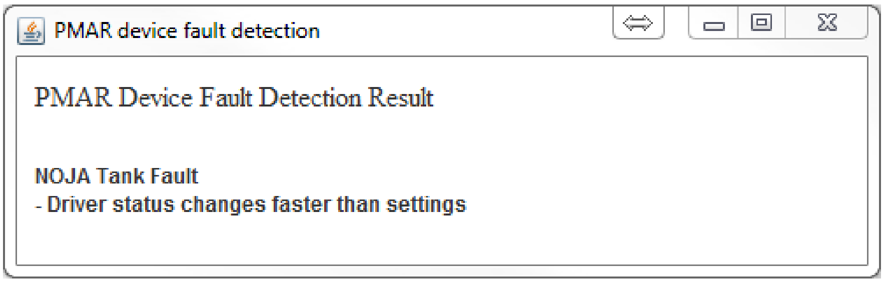


Figure A.8: Identification of tank fault

29/09/2005 01:42:43:290		Open			Manual	
29/09/2005 01:42:43:831		Closed			Undef	
29/09/2005 01:42:44:171		Open			Manual	
29/09/2005 01:42:44:451		Closed			Undef	
29/09/2005 01:42:44:771		Open			Manual	
29/09/2005 01:42:46:072		Closed			Undef	
29/09/2005 01:42:46:232		Open			Manual	
29/09/2005 01:42:46:712		Closed			Undef	
29/09/2005 01:43:00:579		Open			Manual	
29/09/2005 01:43:53:434		Closed			Undef	
29/09/2005 01:43:55:614		Open			Manual	
29/09/2005 01:43:58:155		Closed			Undef	
29/09/2005 21:18:36:417		Open			Manual	
30/09/2005 01:02:33:227		Closed			Undef	

Figure A.9: Validation for tank fault diagnosis

A.2.4 Data for Case Study of Umbilical Cable Fault Diagnosis

With analysis of imported PMAR data, the report states: ***Umbilical Cable fault with the detection of water ingress***, the dialogue box is exhibited in Figure A.10 To validate the automatic diagnosis, Figure 5.12 shows the diagnostic rule and Figure A.11 shows the data fires it.

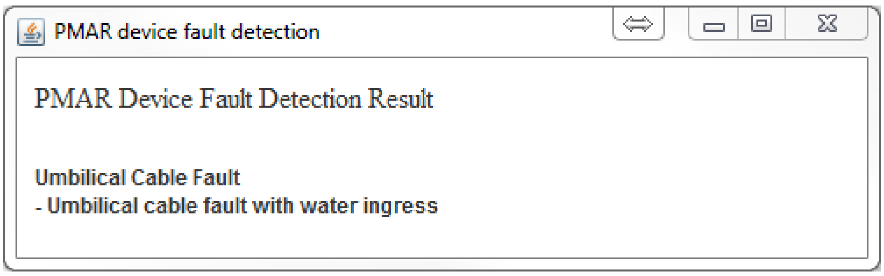


Figure A.10: Identification of umbilical cable fault

22/09/2008 09:32:42:264		Pickup		Start	Uabc>	
22/09/2008 09:34:52:782		Pickup		End	Uabc>	
22/09/2008 09:43:34:543		Pickup		Start	Uabc>	
22/09/2008 09:45:44:927		Pickup		End	Uabc>	
22/09/2008 09:47:55:249		Pickup		Start	Uabc>	
22/09/2008 09:50:05:585		Pickup		End	Uabc>	
22/09/2008 09:52:16:024		Pickup		Start	Uabc>	
22/09/2008 09:54:26:657		Pickup		End	Uabc>	
22/09/2008 09:56:37:139		Pickup		Start	Uabc>	
22/09/2008 09:58:47:190		Pickup		End	Uabc>	
22/09/2008 10:00:57:587		Pickup		Start	Uabc>	
22/09/2008 10:03:07:936		Pickup		End	Uabc>	
22/09/2008 10:07:28:830		Pickup		Start	Uabc>	
22/09/2008 10:09:39:333		Pickup		End	Uabc>	
22/09/2008 10:11:52:007		Pickup		Start	Uabc>	
22/09/2008 10:11:52:015		Pickup		End	Uabc>	
22/09/2008 10:14:00:696		Pickup		Start	Uabc>	
22/09/2008 10:16:10:336		Pickup		End	Uabc>	

Figure A.11: Validation for umbilical cable fault diagnosis

A.3 Source Code for Fault Diagnostic Rules for Chapter 5

The following figures show the source codes of rules to identify PMAR device faults, to categorise overhead line behaviours.

A.4 Source Code for Fault Prognostic Rules for Chapter 6

The following figures show the source code of rules to predict future PMAR operations.

```
rule "Main Processor Module Fault"
    when
        $mpmfilter : device( col1 == "MPM fault", No : indexNo);
        exists device( col1 == "MPM fault", indexNo == No+9 );
        exists device( col1 == "MPM fault", indexNo == No+8 );
        exists device( col1 == "MPM fault", indexNo == No+7 );
        exists device( col1 == "MPM fault", indexNo == No+6 );
        exists device( col1 == "MPM fault", indexNo == No+5 );
        exists device( col1 == "MPM fault", indexNo == No+4 );
        exists device( col1 == "MPM fault", indexNo == No+3 );
        exists device( col1 == "MPM fault", indexNo == No+2 );
        exists device( col1 == "MPM fault", indexNo == No+1 );
        exists device( col1 == "MPM fault", indexNo == No-1 );
        exists device( col1 == "MPM fault", indexNo == No-2 );
        exists device( col1 == "MPM fault", indexNo == No-3 );
        exists device( col1 == "MPM fault", indexNo == No-4 );
        exists device( col1 == "MPM fault", indexNo == No-5 );
        exists device( col1 == "MPM fault", indexNo == No-6 );
        exists device( col1 == "MPM fault", indexNo == No-7 );
        exists device( col1 == "MPM fault", indexNo == No-8 );
        exists device( col1 == "MPM fault", indexNo == No-9 );
        exists device( col1 == "MPM fault", indexNo == No-10 );
    then
    $mpmfilter.setSign1(1);
        msg1.append("Main Processor Module Fault" + No + "---"+ $mpmfilter.getSign1());
        msg2.append("A number of fleeting MPM alarms");
end
```

Figure A.12: The source code of MPM fault diagnosis rule

```
rule "Driver Module Fault"
salience 90
    when
        $mpmfilter : device( col1 == "Excessive To" || col1 == "Excessive Tc"
        || col1 == "Driver not ready" || col1 == "OSM coil SC" || col1 == "OSM coil isolated",
        No : indexNo);
    then
    $mpmfilter.setSign2(1);
        msg1.append("Driver Module Fault" + No + "+++"+ $mpmfilter.getSign2());
        msg2.append("Excessive contact closing/opening result in alarms");
end
```

Figure A.13: The source code of driver module fault diagnosis rule

```
rule "NOJA Tank Fault"
salience 80
    when
        $mpmfilter : device( col1 == "Closed", col3 == "Undef", No : indexNo,TimeV : timeValue);
        exists device( col1 == "Open", indexNo == No-1,timeValue + 10000 > TimeV);
    then
    $mpmfilter.setSign3(1);
        msg1.append("NOJA Tank Fault" + No + "---"+ $mpmfilter.getSign3());
        msg2.append("Driver status changes faster than settings");
end
```

Figure A.14: The source code of tank fault diagnosis rule

```
rule "Microswitch Fault Scenario 1"
salience 100
    when
        $mpmfilter : device3( col1 == "Trip", No : indexNo);
        exists device3( col1 == "Open", indexNo == No+1, col3Int > 60);
    then
        msg5.append("Microswitch Fault 1 - excessive contact opening time");
end

rule "Microswitch Fault Scenario 2"
salience 90
    when
        $mpmfilter : device3( col1 == "Close", No : indexNo);
        exists device3( col1 == "Closed", indexNo == No+1, col3Int > 100);
    then
        msg6.append("Microswitch Fault 2 - excessive contact closing time");
end

rule "Microswitch Fault Scenario 3"
    when
        $mpmfilter : device3( col1 == "Trip", No : indexNo);
        exists device3( col1 == "Open", indexNo == No+1, col3Int < 25);
        exists device3( col1 == "Close", indexNo == No+2);
        exists device3( col1 == "Closed", indexNo == No+3, col3Int < 40);
    then
        msg7.append("Microswitch Fault 3 - unexpected accelerate contact time");
end
```

Figure A.15: The source code of microswitch fault diagnosis rule

```
rule "Umbilical Cable Fault"
salience 70
    when
        $mpmfilter : device( col1 == "Pickup" && col3 == "Uabc>", No : indexNo);
        exists device( col1 == "Pickup" && col3 == "Uabc>", indexNo == No+5 );
        exists device( col1 == "Pickup" && col3 == "Uabc>", indexNo == No+4 );
        exists device( col1 == "Pickup" && col3 == "Uabc>", indexNo == No+3 );
        exists device( col1 == "Pickup" && col3 == "Uabc>", indexNo == No+2 );
        exists device( col1 == "Pickup" && col3 == "Uabc>", indexNo == No+1 );
        exists device( col1 == "Pickup" && col3 == "Uabc>", indexNo == No-1 );
        exists device( col1 == "Pickup" && col3 == "Uabc>", indexNo == No-2 );
        exists device( col1 == "Pickup" && col3 == "Uabc>", indexNo == No-3 );
        exists device( col1 == "Pickup" && col3 == "Uabc>", indexNo == No-4 );
    then
    $mpmfilter.setSign4(1);
        msg1.append("Umbilical Cable Fault" + No + "~~~"+ $mpmfilter.getSign4());
        msg2.append("Umbilical cable fault with water ingress");
end
```

Figure A.16: The source code of umbilical cable fault diagnosis rule

```
rule "Instance: Fault pickups not lead to trips"
salience 100
    when
        $inst : Faultrecord( event_type == "Pickup", No : indexNo);
        exists Faultrecord( event_type == "Pickup", indexNo == No+1);
    then
    $inst.setConseq(Consequence.FAULT_PICKUP);
end

rule "Instance: Single trip"
salience 90
    when
        $inst : Faultrecord( event_type == "Trip", No : indexNo, TimeV : timeValue);
        $instset : Faultrecord( indexNo == No-1);
        not Faultrecord( event_type == "Trip", indexNo == No+2, timeValue < TimeV + 14000);
    then
    $instset.setConseq(Consequence.SINGLE_TRIP);
end

rule "Instance: Multiple trips"
salience 80
    when
        $inst : Faultrecord( event_type == "Trip", No : indexNo, TimeV : timeValue);
        Faultrecord( event_type == "Trip", indexNo == No-2, timeValue + 14000 > TimeV);
        $instset : Faultrecord( indexNo == No-1);
        not Faultrecord( (event_type == "Trip" || == "Lockout"), indexNo == No+2, timeValue < TimeV + 14000);
    then
    $instset.setConseq(Consequence.MULTI_TRIP);
end

rule "Instance: Lockout"
salience 70
    when
        $inst : Faultrecord( event_type == "Lockout", No : indexNo);
        $instset : Faultrecord( indexNo == No-1);
    then
    $instset.setConseq(Consequence.LOCKOUT);
end
```

Figure A.17: The source code of rule to classify PMAR operations

```
rule "General FP group for prediction"
    when
        $inst : Pickuppattern1( trip_index == 1, 8>=total_pickup>=3,
        25<=average_currentamp<=220, 15<=average_pickupdur<=120);
    then
$inst.setGenepickuptotrip(1);
$inst.getEventtotrip();
end
```

Figure A.18: The source code of general predictive rule

```
rule "Specific FP group (EF/SEF) for prediction"
    when
        $inst : Pickuppattern1( trip_index == 1, trip_type ==1; 6>=total_pickup>=3,
        25<=average_currentamp<=80, 15<=average_pickupdur<=100);
    then
$inst.setSpecipickuptotrip(1);
$inst.getEventtotrip();
end
```

Figure A.19: The source code of specific predictive rule

Bibliography

[AB14] A. Abdelmoumene and H. Bentarzi. A review on protective relays developments and trends. *Journal of Energy in Southern Africa*, 25(2):91–95, May 2014. (cited on page 29)

[ABB07] ABB. Distribution Automation Handbook: Automatic Reclosing. Technical report, 2007. (cited on pages viii, 32, 34, 36)

[Abb08] O. A. Abbas. Comparisons between data clustering algorithm. *The International Arab Journal of Information Technology*, 5(3):320–325, Jul 2008. (cited on page 136)

[ABKS99] Mihael Ankerst, Markus M Breunig, Hans-Peter Kriegel, and Jörg Sander. Optics: ordering points to identify the clustering structure. In *Management of Data Conference*, volume 28, pages 49–60, 1999. (cited on pages x, 144, 145)

[Act15] ActewAGL Distribution. Distribution Annual Planning Report 2015. Technical report, 2015. (cited on pages viii, 12)

[Agu14] Aguas De Murcia. Smart water - Green Digital Charter, 2014. URL: `http://www.greendigitalcharter.eu/wp-content/uploads/2014/06/Emuasa-Smartwater_W.pdf`. (cited on page 24)

[AHF+98] T. Aschwanden, M. Hassig, J. Fuhr, P. Lorin, V. Der Houhanessian, W. Zaengl, A. Schenk, P. Zweiacker, and A. Piras. Development and application of new condition assessment methods for power transformers. In *International Conference on Large High Voltage Electric Systems. CIGRE'98. Session Papers*, pages 1–10, 1998. (cited on page 67)

[Ale02] R. L. Alexander. Intelligent electronic device (ied) technology scada and 3ø metering. In *Rural Electric Power Conference, 2002. 2002 IEEE*, pages C6/1–C6/3, 2002. doi:10.1109/.2002.1002298. (cited on page 19)

[Alp10] E. Alpaydn. *Introduction to Machine Learning*. The MIT Press, 2010. (cited on pages 52, 57, 136)

[AMG+97] D. Atanackovic, D. T. McGillis, F. D. Galiana, J. Cheng, and L. Loud. An integrated knowledge based model for power system planning. *IEEE Expert*, 12(4):65–71, Jul 1997. doi:10.1109/64.608198. (cited on page 6)

[AONI12] I. Ahmed, S. Obermeier, M. Naedele, and G. G. Richard III. Scada systems: Challenges for forensic investigators. *Computer*, 45(12):44–51, Dec 2012. doi:10.1109/MC.2012.325. (cited on pages viii, 19, 20)

[AP94] A. Aamodt and E. Plaza. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, 7(1):39–59, 1994. (cited on pages ix, 70)

[APP14] APPA. Evaluation of Data Submitted in APPA's 2013 Distribution System Reliability & Operations Survey. Technical report, APPA, 2014. (cited on page 2)

[AS97] R. Aggarwal and Y. Song. Artificial neural networks in power systems I: General introduction to neural computing. *Power Engineering Journal*, 11(3):129–134, June 1997. doi:10.1049/pe:19970306. (cited on page 53)

[AS09] M. M. Ahmed and W. L. Soo. Customized scada system for low voltage distribution automation system. In *2009 Transmission Distribution Conference Exposition: Asia and Pacific*, pages 1–4, Oct 2009. doi:10.1109/TD-ASIA.2009.5357001. (cited on page 24)

[AS10] R. Akerkar and P. Sajja. *Knowledge-Based Systems*. Jones & Bartlett Learning, LLC, 2010. URL: `http://www.tmrfindia.org/eseries/ebookv1-c1.pdf`. (cited on pages ix, 61, 61)

[BA13] R. Billinton and R. N. Allan. *Reliability Evaluation of Power Systems*. Springer Science & Business Media, 2013. (cited on page 36)

[Bau10] Eric Bauer. *Reliability and Availability Concepts*, pages 0–. Wiley-IEEE Press, 2010. URL: `http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5733168`, `doi:10.1002/9781118075104.ch1`. (cited on page 36)

[BDF$^+$93] A. Beschta, O. Dressler, H. Freitag, M. Montag, and P. Struss. A model-based approach to fault localisation in power transmission networks. *Intelligent Systems Engineering*, 2(1):3–14, Spring 1993. (cited on page 67)

[BH12] C. Bayliss and B. Hardy. *Transmission and Distribution Electrical Engineering*. Newnes, 2012. (cited on page 13)

[BI98] P. Burrell and D. Inman. An expert system for the analysis of faults in electricity supply networks: Problems and achievements. *Computers in Industry*, 37(2):113–123, Sep 1998. `doi:10.1016/S0166-3615(98)00087-6`. (cited on page 44)

[Bla13] S. Blair. *The Analysis and Application of Resistive Superconducting Fault Current Limiters in Present and Future Power Systems*. PhD thesis, University of Strathclyde, 2013. (cited on page 28)

[Blu07] Steven W. Blume. *System Protection*, pages 161–178. Wiley-IEEE Press, 2007. URL: `http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5396682`, `doi:10.1002/9780470185810.ch7`. (cited on page 26)

[BMK$^+$95] G. M. Burt, J. R. McDonald, A. G. King, J. Spiller, D. Brooke, and R. Samwell. Intelligent on-line decision support for distribution system control and operation. *IEEE Transactions on Power Systems*, 10(4):1820–1827, Nov 1995. doi:10.1109/59.476046. (cited on page 44)

[BMM+98] S. C. Bell, S. D. J. McArthur, J. R. McDonald, G. M. Burt, R. Mather, and T. Cumming. Model-based analysis of protection system performance. *IEE Proceedings - Generation, Transmission and Distribution*, 145(5):547–552, Sep 1998. doi:10.1049/ip-gtd:19982183. (cited on pages 3, 6, 44, 67)

[Bra01] I. Bratko. *Prolog Programming for Artificial Intelligence: programming for artificial intelligence.* Pearson Education, 2001. (cited on page 63)

[Bri11] British Power International. Quality of Service Incentive Scheme Audit of Interruptions Reporting 2009/10 Final Report. Technical report, 2011. (cited on page 40)

[BS14] M. Bianchini and F. Scarselli. On the complexity of neural network classifiers: A comparison between shallow and deep architectures. *Neural Networks and Learning Systems, IEEE Transactions on*, 25(8):1553–1565, Aug 2014. doi:10.1109/TNNLS.2013.2293637. (cited on page 54)

[Bus13] Stephen F. Bush. *Distribution*, pages 576–. Wiley-IEEE Press, 2013. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6740208, doi:10.1002/9781118820216.ch4. (cited on page 15)

[BWM14] Hassan Bevrani, Masayuki Watanabe, and Yasunori Mitani. *Power System Control: Fundamentals and New Perspectives*, pages 70–95. Wiley-IEEE Press, 2014. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6837332, doi:10.1002/9781118852422.ch05. (cited on page 2)

[CBG14] S. Charbonnier, N. Bouchair, and P. Gayet. Analysis of fault diagnosability from scada alarms signatures using relevance indices. In

2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pages 2739–2744, Oct 2014. doi:10.1109/SMC.2014.6974342. (cited on pages 24, 43)

[CdVR07] J. Coser, D. T. do Vale, and J. G. Rolim. Design and training of artificial neural networks for locating low current faults in distribution systems. In *Intelligent Systems Applications to Power Systems, 2007. ISAP 2007. International Conference on*, pages 1–6, Nov 2007. doi:10.1109/ISAP.2007.4441599. (cited on page 55)

[Che11] W. H. Chen. Fault section estimation using fuzzy matrix-based reasoning methods. *IEEE Transactions on Power Delivery*, 26(1):205–213, Jan 2011. doi:10.1109/TPWRD.2010.2061873. (cited on page 3)

[Che12] W. H. Chen. Decentralized fault diagnosis and its hardware implementation for distribution substations. *IEEE Transactions on Power Delivery*, 27(2):902–909, April 2012. doi:10.1109/TPWRD.2011.2179949. (cited on pages 3, 43)

[CJ98] D. Coury and D. Jorge. Artificial neural network approach to distance protection of transmission lines. *Power Delivery, IEEE Transactions on*, 13(1):102–108, Jan 1998. doi:10.1109/61.660861. (cited on page 55)

[CK09] Ali Chowdhury and Don Koval. *Reliability Principles*, pages 45–77. Wiley-IEEE Press, 2009. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5396890, doi:10.1002/9780470459355.ch3. (cited on page 36)

[Cla11] Allen L. Clapp. *Circuit Breakers, Reclosers, Switches, and Fuses*, pages 203–208. Wiley-IEEE Standards Association, 2011. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5989331, doi:10.1109/9781118187890.ch17. (cited on page 31)

[CQF+11] B. Chen, Y. N. Qiu, Y. Feng, P. J. Tavner, and W. W. Song. Wind turbine scada alarm pattern recognition. In *Renewable Power Generation (RPG 2011), IET Conference on*, pages 1–6, Sept 2011. doi:10.1049/cp.2011.0164. (cited on page 24)

[CS15] V. Catterson and B. Sheng. Deep neural networks for understanding and diagnosing partial discharge data. In *Electrical Insulation Conference (EIC), 2015 IEEE*, pages 1–6, June 2015. (cited on page 54)

[Das12] J. C. Das. *Protective Relaying*, pages 203–265. Wiley-IEEE Press, 2012. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6307940, doi:10.1002/9781118402498.ch7. (cited on page 31)

[DMA+15] R. Das, V. Madani, F. Aminifar, J. McDonald, S. S. Venkata, D. Novosel, A. Bose, and M. Shahidehpour. Distribution automation strategies: Evolution of technologies and the business case. *IEEE Transactions on Smart Grid*, 6(4):2166–2175, July 2015. doi:10.1109/TSG.2014.2368393. (cited on page 2)

[DMM03] E. Davidson, S. McArthur, and J. McDonald. A toolset for applying model-based reasoning techniques to diagnostics for power systems protection. *Power Systems, IEEE Transactions on*, 18(2):680–687, May 2003. doi:10.1109/TPWRS.2003.810981. (cited on pages ix, 66)

[DMM+06] E. M. Davidson, S. D. J. McArthur, J. R. McDonald, T. Cumming, and I. Watt. Applying multi-agent system technology in practice: automated management and analysis of scada and digital fault recorder data. *IEEE Transactions on Power Systems*, 21(2):559–567, May 2006. doi:10.1109/TPWRS.2006.873109. (cited on pages 3, 45)

[DWZ+15] R. C. Duan, F. H. Wang, J. Zhang, R. H. Huang, and X. Zhang. Data mining & pattern recognition of voltage sag based on k-means clustering algorithm. In *2015 IEEE Power Energy Society General Meeting*, pages 1–5, July 2015. doi:10.1109/PESGM.2015.7286079. (cited on page 60)

[DZG13] T. Dekkers, M. Zut, and K. Gleeson. A new approach to intermittant fault management in low voltage networks to reduce the number of repeat interruptions. In *Electricity Distribution (CIRED 2013), 22nd International Conference and Exhibition on*, pages 1–3, June 2013. doi:10.1049/cp.2013.1007. (cited on pages 3, 5, 46)

[EAJMB05] K. El-Arroudi, G. Joos, D. McGillis, and R. Brearley. Comprehensive transmission distance protection settings using an intelligent-based analysis of events and consequences. *Power Delivery, IEEE Transactions on*, 20(3):1817–1824, July 2005. doi:10.1109/TPWRD.2005.848656. (cited on page 63)

[EAT14] EATON. Sectionalizers: Catalog information, 2014. URL: http://www.eaton.com/ecm/idcplg?IdcService=GET_FILE&allowInterrupt=1&RevisionSelectionMethod=LatestReleased&Rendition=Primary&dDocName=PCT_1130878. (cited on page 31)

[Ecl14] Eclipse Foundation. Eclipse, 2014. URL: https://eclipse.org/. (cited on page 93)

[EDM+13] 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 Transactions on Power Systems*, 28(3):2584–2591, Aug 2013. doi:10.1109/TPWRS.2013.2243763. (cited on pages 45, 67)

[EH95] Mohamed E. El-Hawary. *Faults on Electric Energy Systems*, pages 469–540. Wiley-IEEE Press, 1995. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5311843, doi:10.1109/9780470544464.ch8. (cited on page 27)

[EH08] Mohamed E. El-Hawary. *Faults and Protection of Electric Energy Systems*, pages 400–. Wiley-IEEE Press, 2008. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5397121, doi:10.1002/9780470411377.ch7. (cited on page 26)

[ES13] Mircea Eremia and Mohammad Shahidehpour. *Power System Protection*, pages 737–785. Wiley-IEEE Press, 2013. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6482750, doi:10.1002/9781118516072.ch12. (cited on page 26)

[Far10] H. Farhangi. The path of the smart grid. *IEEE Power and Energy Magazine*, 8(1):18–28, January 2010. doi:10.1109/MPE.2009.934876. (cited on page 1)

[Feh16] Ralph Fehr. *Protective Device Selection And Coordination*, pages 440–. Wiley-IEEE Press, 2016. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7394671, doi:10.1002/9781119065180.ch5. (cited on page 31)

[FK86] C. Fukui and J. Kawakami. An expert system for fault section estimation using information from protective relays and circuit breakers. *Power Delivery, IEEE Transactions on*, 1(4):83–90, Oct 1986. doi:10.1109/TPWRD.1986.4308033. (cited on page 64)

[Fog06] David B. Fogel. *Defining Artificial Intelligence*, pages 1–32. Wiley-IEEE Press, 2006. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5237936, doi:10.1002/0471749214.ch1. (cited on page 50)

[For82] C. L. Forgy. Rete: A fast algorithm for the many pattern/many object pattern match problem. *Artificial Intelligence*, 19(1):17–37, 1982. doi:10.1016/0004-3702(82)90020-. (cited on page 93)

[FTKM12] A. Filippo, Jr. A. Torres, B. Kjerfve, and A. Monat. Application of artificial neural network (ANN) to improve forecasting of sea level. *Ocean and Coastal Management*, 55(0):101–110, 2012. URL: http://www.sciencedirect.com/science/article/pii/S0964569111001475, doi:http://dx.doi.org/10.1016/j.ocecoaman.2011.09.007. (cited on page 54)

[GA11] C. Grosan and A. Abraham. *Intelligent Systems: A Modern Approach*. Intelligent Systems Reference Library. Springer Berlin Heidelberg, 2011. (cited on pages 62, 63)

[GASD11] R. Greer, W. Allen, J. Schnegg, and A. Dulmage. Distribution automation systems with advanced features. In *Rural Electric Power Conference (REPC), 2011 IEEE*, pages C4–1–C4–15, April 2011. doi:10.1109/REPCON.2011.5756721. (cited on page 16)

[GB10] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1–6, 2010. (cited on page 54)

[GCF16] M. J. Ghorbani, M. A. Choudhry, and A. Feliachi. A multiagent design for power distribution systems automation. *IEEE Transactions on Smart Grid*, 7(1):329–339, Jan 2016. doi:10.1109/TSG.2015.2453884. (cited on pages 3, 32)

[GD87] D. J. Gaushell and H. T. Darlington. Supervisory control and data acquisition. *Proceedings of the IEEE*, 75(12):1645–1658, Dec 1987. doi:10.1109/PROC.1987.13932. (cited on page 19)

[gJmH12] D. g. Ji and D. m. Huang. A research based on k-means clustering and artificial fish-swarm algorithm for the vehicle routing optimization. In *Natural Computation (ICNC), 2012 Eighth International Conference on*, pages 1141–1145, May 2012. doi:10.1109/ICNC.2012.6234729. (cited on page 59)

[GK09] Y. Guan and M. Kezunovic. Implementing intelligent techniques for the advanced alarm processing. In *Intelligent System Applications to Power Systems, 2009. ISAP '09. 15th International Conference on*, pages 1–6, Nov 2009. doi:10.1109/ISAP.2009.5352909. (cited on page 47)

[GP13] T. Gu and F. Provoost. Improving network performance by recognition and location of self-extinguishing faults. In *Electricity Distribu-*

tion (CIRED 2013), 22nd International Conference and Exhibition on, pages 1–4, June 2013. doi:10.1049/cp.2013.0728. (cited on page 5)

[GR13] M. Gillick and J. Rodgers. Management of scottish and southern energys low voltage distribution cable network using smart fuses with online fault detection. In *Network Components (CIRED 2013), 22nd International Conference and Exhibition on*, pages 1–2, June 2013. (cited on pages 5, 46, 48)

[Gri11] Alstom Grid. *Network protection & automation guide*. Alstom Grid, 2011. (cited on pages viii, 30, 31)

[GSO12] J. Glover, M. Sarma, and T. Overbye. *Power System Analysis and Design*. Cengage Learning, 2012. (cited on page 26)

[Hat15] B Hatton. Engineering Design Standard, 2015. URL: https://library.ukpowernetworks.co.uk/library/en/g81/Design_and_Planning/Planning_and_Design/Documents/EDS+08-0136+LV+Network+Design+Standard.pdf. (cited on page 13)

[HHC+90] Ku-Long Ho, Yuan-Yih Hsu, Chuan-Fu Chen, Tzong-En Lee, Chih-Chien Liang, Tsau-Shin Lai, and Kung-Keng Chen. Short term load forecasting of taiwan power system using a knowledge-based expert system. *IEEE Transactions on Power Systems*, 5(4):1214–1221, Nov 1990. doi:10.1109/59.99372. (cited on page 6)

[HLL15] S. C. Huang, C. N. Lu, and Y. L. Lo. Evaluation of ami and scada data synergy for distribution feeder modeling. *IEEE Transactions on Smart Grid*, 6(4):1639–1647, July 2015. doi:10.1109/TSG.2015.2408111. (cited on page 42)

[HMMM03] J. A. Hossack, J. Menal, S. D. J. McArthur, and J. R. McDonald. A multiagent architecture for protection engineering diagnostic assistance. *IEEE Transactions on Power Systems*, 18(2):639–647, May 2003. doi:10.1109/TPWRS.2003.810910. (cited on pages 3, 44)

[Hon15] Q. Hong. *Design of an Intelligent System for Validation of Protection Settings*. PhD thesis, University of Strathclyde, 2015. (cited on pages viii, 27)

[HPF12] R. A. Hooshmand, M. Parastegari, and Z. Forghani. Adaptive neuro-fuzzy inference system approach for simultaneous diagnosis of the type and location of faults in power transformers. *IEEE Electrical Insulation Magazine*, 28(5):32–42, September 2012. doi:10.1109/MEI.2012.6268440. (cited on page 64)

[IA09] A. Ipakchi and F. Albuyeh. Grid of the future. *IEEE Power and Energy Magazine*, 7(2):52–62, March 2009. doi:10.1109/MPE.2008.931384. (cited on page 1)

[Ibr12] Mohamed A. Ibrahim. *Power System Disturbance Analysis Function*, pages 1–32. Wiley-IEEE Press, 2012. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6047706, doi:10.1002/9781118172094.ch1. (cited on page 2)

[IEC07] IEC International Electrotechnical Commission. Efficient Electrical Energy Transmission and Distribution, 2007. URL: http://web.vdw.co.za/Portals/15/Documents/Presentations/transmission.pdf. (cited on page 13)

[IEE93] Ieee recommended practice for master/remote supervisory control and data acquisition (scada) communications. *IEEE Std 999-1992*, 1993. doi:10.1109/IEEESTD.1993.114462. (cited on page 19)

[IEE06] IEEE. IEEE recommended practice for excitation system models for power system stability studies. *IEEE Std 421.5-2005 (Revision of IEEE Std 421.5-1992)*, 2006. doi:10.1109/IEEESTD.2006.99499. (cited on page 26)

[IEE07] IEEE Smart Distribution Working Group. Smart Grid for Distribution Systems: The Benefits and Challenges of Distribution Automation (DA) (Draft Version 2) White Paper for NIST. Technical report, 2007. (cited on page 17)

[IEE12] Ieee guide for electric power distribution reliability indices - redline. *IEEE Std 1366-2012 (Revision of IEEE Std 1366-2003) - Redline*, pages 1–92, May 2012. (cited on page 38)

[IRS07] J. Izykowski, E. Rosolowski, and M. M. Saha. Postfault analysis of operation of distance protective relays of power transmission lines. *IEEE Transactions on Power Delivery*, 22(1):74–81, Jan 2007. `doi:10.1109/TPWRD.2006.877084`. (cited on page 5)

[JBRRGM14] M. Jaya Bharata Reddy, D. Rajesh, P. Gopakumar, and D. Mohanta. Smart fault location for smart grid operation using RTUs and computational intelligence techniques. *Systems Journal, IEEE*, 8(4):1260–1271, Dec 2014. `doi:10.1109/JSYST.2014.2303833`. (cited on page 55)

[KC11] S. Karnouskos and A. W. Colombo. Architecting the next generation of service-based scada/dcs system of systems. In *IECON 2011 - 37th Annual Conference on IEEE Industrial Electronics Society*, pages 359–364, Nov 2011. `doi:10.1109/IECON.2011.6119279`. (cited on page 19)

[KD11] S. Khalid and B. Dwivedi. Power quality issues, problems, standards & their effects in industry with corrective means. *International Journal of Advances in Engineering & Technology*, 1(2):1–11, May 2011. (cited on page 17)

[KDS10] Rajeev Kumar, M. L. Dewal, and Kalpana Saini. Utility of scada in power generation and distribution system. In *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on*, volume 6, pages 648–652, July 2010. `doi:10.1109/ICCSIT.2010.5564689`. (cited on page 19)

[KEDH02a] M. Kandil, S. El-Debeiky, and N. Hasanien. Long-term load forecasting for fast developing utility using a knowledge-based expert system. *Power Systems, IEEE Transactions on*, 17(2):491–496, May 2002. `doi:10.1109/TPWRS.2002.1007923`. (cited on page 63)

[KEDH02b] M. S. Kandil, S. M. El-Debeiky, and N. E. Hasanien. Long-term load forecasting for fast developing utility using a knowledge-based expert system. *IEEE Transactions on Power Systems*, 17(2):491–496, May 2002. doi:10.1109/TPWRS.2002.1007923. (cited on page 6)

[Ker01] W. H. Kersting. *Distribution system modeling and analysis.* CRC Press, 2001. (cited on page 13)

[Kez11] M. Kezunovic. Translational knowledge: From collecting data to making decisions in a smart grid. *Proceedings of the IEEE*, 99(6):977–997, June 2011. doi:10.1109/JPROC.2011.2112631. (cited on pages 47, 48)

[KG09] M. Kezunovic and Yufan Guan. Intelligent alarm processing: From data intensive to information rich. In *System Sciences, 2009. HICSS '09. 42nd Hawaii International Conference on*, pages 1–8, Jan 2009. doi:10.1109/HICSS.2009.268. (cited on page 46)

[KHML13] M. Koivisto, P. Heine, I. Mellin, and M. Lehtonen. Clustering of connection points and load modeling in distribution systems. *IEEE Transactions on Power Systems*, 28(2):1255–1265, May 2013. doi:10.1109/TPWRS.2012.2223240. (cited on page 60)

[Kir14] Kirti. Scada: Supervisory control and data acquisition. *International Journal Of Engineering And Computer Science*, 3(1):3743–3751, Jan 2014. (cited on page 22)

[KM10] M. Kim and J. J. Metzner. A key exchange method for intelligent electronic devices in distribution automation. *IEEE Transactions on Power Delivery*, 25(3):1458–1464, July 2010. doi:10.1109/TPWRD.2010.2040490. (cited on page 2)

[KNUF92] T. Kimura, S. Nishimatsu, Y. Ueki, and Y. Fukuyama. Development of an expert system for estimating fault section in control center based on protective system simulation. *Power Delivery, IEEE Transactions on*, 7(1):167–172, Jan 1992. doi:10.1109/61.108904. (cited on page 64)

[KPS+10] M. Kezunovic, T. Popovic, S. Sternfeld, B. Fardanesh, B. Clowe, and P. Myrda. Substation automated data analysis: Deployment challenges. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 137–142, Oct 2010. `doi:10.1109/SMARTGRID.2010.5622031`. (cited on page 46)

[KRFS94] M. Kezunovic, I. Rikalo, C. W. Fromen, and D. R. Sevcik. Expert system reasoning streamlines disturbance analysis. *IEEE Computer Applications in Power*, 7(2):15–19, April 1994. `doi:10.1109/67.273781`. (cited on page 5)

[KSE+00] W. A. W. Kamarulzaman, A. M. Sufian, R. Embong, M. Shafik, and M. Taha. Substation alarm monitoring system. In *TENCON 2000. Proceedings*, volume 2, pages 603–608 vol.2, 2000. `doi:10.1109/TENCON.2000.888808`. (cited on page 24)

[Kum10] Recent advances in scada alarm system. *International Journal of Smart Home*, 4(4):1–10, October 2010. (cited on page 19)

[KV91] D. Kirschen and T. Volkmann. Guiding a power system restoration with an expert system. *Power Systems, IEEE Transactions on*, 6(2):558–566, May 1991. `doi:10.1109/59.76698`. (cited on pages 6, 63)

[KW87] J. Kleer and B. Williams. Diagnosing Multiple Faults, 1987. URL: `http://groups.csail.mit.edu/mers/papers/kle-wil-92.pdf`. (cited on page 68)

[LFST94] R. Leitch, H. Freitag, A. Stefanini, and G. Tornielli. Using the artist approach for diagnosing power transmission networks. *Intelligent Systems Engineering*, 3(3):125–137, Autumn 1994. (cited on page 67)

[Lig06] A. Ligeza. *Logical Foundations for Rule-Based Systems*. Studies in Computational Intelligence. Springer Berlin Heidelberg, 2006. (cited on page 62)

[LK05] X. Luo and M. Kezunovic. An expert system for diagnosis of digital relay operation. In *Proceedings of the 13th International Conference* *on, Intelligent Systems Application to Power Systems*, pages 1–6, Nov 2005. doi:10.1109/ISAP.2005.1599258. (cited on page 5)

[LL96] S. Lee and C. Liu. Intelligent approach to coordination identification in distance relaying. In *Intelligent Systems Applications to Power Systems, 1996. Proceedings, ISAP 1996., International Conference on*, pages 62–67, Jan 1996. doi:10.1109/ISAP.1996.501045. (cited on page 63)

[LPS+13] A. LEITAO, C. M. Pinto, V. SANTOS, P. DIAS, P. VIEGAS, D. MARSH, and P. BIZARRO. Smart alarm processing. In *CIRED - 22nd International Conference on Electricity Distribution*, June 2013. doi:10.1049/cp.2013.0973. (cited on page 24)

[Lug09] G. Luger. *Artificial intelligence : structures and strategies for complex problem solving.* Pearson Education, 2009. (cited on pages viii, 51, 52, 54, 60, 62, 63, 65, 66, 68, 69, 70, 92)

[Lun11] B. Lundqvist. 100 years of relay protection, the Swedish ABB relay history, 2011. URL: https://abblibrary.s3.amazonaws.com/public/c1256d32004634bac1256e19006fd705/PAPER_2001_08_en_100_Years_of_Relay_Protection__the_Swedish_ABB_Relay_History.pdf. (cited on page 29)

[LYYJ90] S. Lee, S. Yoon, M. Yoon, and J. Jang. An expert system for protective relay setting of transmission systems. *Power Delivery, IEEE Transactions on*, 5(2):1202–1208, Apr 1990. doi:10.1109/61.53142. (cited on page 63)

[M. 94] M. Enns, L. Budler, T. Cease et al. Potential applications of expert systems to power system protection. *Power Delivery, IEEE Transactions on*, 9(2):720–728, Apr 1994. doi:10.1109/61.296249. (cited on pages 51, 63)

[Mac67] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1:*

Statistics, pages 281–297, Berkeley, Calif., 1967. University of California Press. URL: `http://projecteuclid.org/euclid.bsmsp/1200512992`. (cited on page 59)

[MBM+98] S. D. J. McArthur, S. C. Bell, J. R. McDonald, R. Mather, and T. Cumming. The development of an advanced suite of data interpretation facilities for the analysis of power system disturbances, 1998. (cited on pages 44, 48)

[MC11] F. Murtagh and P. Contreras. Algorithms for hierarchical clustering: an overview. *The Jounral of WIREs: Data mining and knowledge discovery*, 2(1):86–97, Dec 2011. `doi:10.1002/widm.53`. (cited on page 136)

[McD93] J. D. McDonald. Developing and defining basic scada system concepts. In *Rural Electric Power Conference, 1993. Papers Presented at the 37th Annual Conference*, pages B3/1–B3/5, Apr 1993. `doi:10.1109/REPCON.1993.239563`. (cited on page 19)

[McD03] J. D. McDonald. Substation automation. ied integration and availability of information. *IEEE Power and Energy Magazine*, 1(2):22–31, Mar 2003. `doi:10.1109/MPAE.2003.1192023`. (cited on page 3)

[MD06] S. D. J. McArthur and E. M. Davidson. Automated post-fault diagnosis of power system disturbances. In *2006 IEEE Power Engineering Society General Meeting*, pages 6 pp.–, 2006. `doi:10.1109/PES.2006.1709651`. (cited on page 45)

[MDA+15] 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 Transactions on Smart Grid*, 6(4):2157–2165, July 2015. `doi:10.1109/TSG.2014.2368382`. (cited on pages 1, 2)

[MDC+07a] S. D. J. McArthur, E. M. Davidson, V. M. Catterson, A. L. Dimeas, N. D. Hatziargyriou, F. Ponci, and T. Funabashi. Multi-agent systems for power engineering applications - part i: Concepts, approaches, and technical challenges. *IEEE Transactions on Power Systems*, 22(4):1743–1752, Nov 2007. doi:10.1109/TPWRS.2007.908471. (cited on page 45)

[MDC+07b] S. D. J. McArthur, E. M. Davidson, V. M. Catterson, A. L. Dimeas, N. D. Hatziargyriou, F. Ponci, and T. Funabashi. Multi-agent systems for power engineering applications - part ii: Technologies, standards, and tools for building multi-agent systems. *IEEE Transactions on Power Systems*, 22(4):1753–1759, Nov 2007. doi:10.1109/TPWRS.2007.908472. (cited on page 45)

[MDDM03] S. McArthur, E. Davidson, G. Dudgeon, and J. McDonald. Toward a model integration methodology for advanced applications in power engineering. *Power Systems, IEEE Transactions on*, 18(3):1205–1206, Aug 2003. doi:10.1109/TPWRS.2003.814908. (cited on page 66)

[MDM+96] 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. *Power Delivery, IEEE Transactions on*, 11(4):1748–1754, Oct 1996. doi:10.1109/61.544253. (cited on pages 6, 67, 68)

[MH08] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008. (cited on pages x, 137, 138)

[MIK+95] T. Minakawa, Y. Ichikawa, M. Kunugi, K. Shimada, N. Wada, and M. Utsunomiya. Development and implementation of a power system fault diagnosis expert system. *Power Systems, IEEE Transactions on*, 10(2):932–940, May 1995. doi:10.1109/59.387936. (cited on page 64)

[Mit97] T. Mitchell. *Machine Learning*. McGraw-Hill Science/Engineering/Math, 1997. (cited on pages viii, 52, 53, 54)

[MM05] V. K. Mehta and R. Mehta. *Principles of Power System: Including Generation, Transmission, Distribution, Switchgear and Protection: for B.E/B.Tech., AMIE and Other Engineering Examinations.* S. Chand, 2005. (cited on page 13)

[MMBB95] S. D. J. MacArthur, J. R. McDonald, S. C. Bell, and G. M. Burt. Expert systems and model based reasoning for protection performance analysis. In *Artificial Intelligence Applications in Power Systems, IEE Colloquium on*, pages 1–4, Apr 1995. doi:10.1049/ic:19950493. (cited on page 5)

[Mom07] J. A. Momoh. *Electric power distribution, automation, protection, and control.* CRC Press, Taylor & Francis Group, 2007. (cited on pages viii, 11, 14, 16, 18, 31, 34, 38)

[MR15] M. Moreto and J. G. Rolim. Using disturbance records to automate the diagnosis of faults and operational procedures in power generators. *IET Generation, Transmission Distribution*, 9(15):2389–2397, 2015. doi:10.1049/iet-gtd.2014.0785. (cited on page 3)

[MRJ97] H. Monsef, A. M. Ranjbar, and S. Jadid. Fuzzy rule-based expert system for power system fault diagnosis. *IEE Proceedings - Generation, Transmission and Distribution*, 144(2):186–192, Mar 1997. doi:10.1049/ip-gtd:19970799. (cited on page 64)

[MSAA15] S. S. Mousavi-Seyedi, F. Aminifar, and S. Afsharnia. Parameter estimation of multiterminal transmission lines using joint pmu and scada data. *IEEE Transactions on Power Delivery*, 30(3):1077–1085, June 2015. doi:10.1109/TPWRD.2014.2369500. (cited on page 42)

[MSJ04] S. D. J. McArthur, S. M. Strachan, and G. Jahn. The design of a multi-agent transformer condition monitoring system. *IEEE Transactions on Power Systems*, 19(4):1845–1852, Nov 2004. doi:10.1109/TPWRS.2004.835667. (cited on page 60)

[NDMM09] G. Napier, E. M. Davidson, S. D. J. McArthur, and M. J. R. McDonald. An automated fault analysis system for sp energy networks: Requirements, design and implementation. In *2009 IEEE Power Energy Society General Meeting*, pages 1–7, July 2009. doi:10.1109/PES.2009.5275427. (cited on page 45)

[NGW07] J. Northcote-Green and R. Wilson. *Control and Automation of Electrical Power Distribution Systems.* CRC Press, Taylor & Francis Group, 2007. (cited on pages viii, 11, 16, 17, 36, 38)

[NKF+16] K. Neshatpour, A. Koohi, F. Farahmand, R. Joshi, S. Rafatirad, A. Sasan, and H. Homayoun. Big biomedical image processing hardware acceleration: A case study for k-means and image filtering. In *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1134–1137, May 2016. doi:10.1109/ISCAS.2016.7527445. (cited on page 59)

[NOJ16] NOJA Power. OSM: Product Guide, 2016. URL: http://www.spenergynetworks.co.uk/userfiles/file/ESDD-02-011_Issue_4_DM_Section_16b_Application_of_Overhead_Line_Switchgear_Protection_Systems.pdf. (cited on pages viii, 32, 33)

[NVI16] NVIDIA. Deep learning on gpus, 2016. URL: http://on-demand.gputechconf.com/gtc/2015/webinar/deep-learning-course/intro-to-deep-learning.pdf. (cited on page 54)

[NVKH11] A. Nolde, H. Vennegeerts, S. Krahl, and HJ. Haubrich. Distribution of reliability indices in electrical networks. In *CIRED 21st Int. Conf*, pages 1–4, June 2011. (cited on page 2)

[OAA+10] M. L. Othman, I. Aris, S. M. Abdullah, M. L. Ali, and M. R. Othman. Knowledge discovery in distance relay event report: A comparative data-mining strategy of rough set theory with decision tree. *IEEE Transactions on Power Delivery*, 25(4):2264–2287, Oct 2010. doi:10.1109/TPWRD.2010.2055587. (cited on page 6)

[Off04] Office of the Manager National Communications System. Supervisory Control and Data Acquisition (SCADA) Systems. Technical report, NATIONAL COMMUNICATIONS SYSTEM, 2004. (cited on pages 2, 19, 19, 21)

[ofg02] ofgem. Electricity Distribution Annual Report for 2010-11. Technical report, 2002. (cited on page 40)

[ofg05] ofgem. Quality of Service Regulatory Instructions and Guidance version 5. Technical report, 2005. (cited on page 40)

[ofg09] ofgem. RIIO-ED1 Glossary of terms. Technical report, 2009. (cited on page 40)

[Ora14] Oracle. Java, 2014. URL: `http://java.com/en/`. (cited on page 84)

[Ora17] Oracle. MySQL, 2017. URL: `https://www.mysql.com/`. (cited on page 84)

[OTYB12] O. Ozgonenel, D. W. P. Thomas, T. Yalcin, and I. N. Bertizlioglu. Detection of blackouts by using k-means clustering in a power system. In *Developments in Power Systems Protection, 2012. DPSP 2012. 11th International Conference on*, pages 1–6, April 2012. `doi:10.1049/cp.2012.0079`. (cited on page 60)

[PKK15] T. Popovic, M. Kezunovic, and B. Krstajic. Smart grid data analytics for digital protective relay event recordings. *Information Systems Frontiers*, 17(3):591–600, June 2015. `doi:10.1007/s10796-013-9434-9`. (cited on pages 2, 3, 5, 47, 48)

[PL97] Y. Park and K. Lee. Application of expert system to power system restoration in sub-control center. *Power Systems, IEEE Transactions on*, 12(2):629–635, May 1997. `doi:10.1109/59.589628`. (cited on pages 6, 63)

[PW03] B. Peischl and F. Wotawa. Model-based diagnosis or reasoning from first principles. *Intelligent Systems, IEEE*, 18(3):32–37, May 2003. `doi:10.1109/MIS.2003.1200725`. (cited on pages 65, 66)

[PWN92] M. Pfau-Wagenbauer and W. Nejdl. Integrating model-based and heuristic features in a real-time expert system for power distribution networks. In *Artificial Intelligence for Applications, 1992., Proceedings of the Eighth Conference on*, pages 303–309, Mar 1992. doi:10.1109/CAIA.1992.200046. (cited on page 6)

[QFS+16] Y. Qiu, Y. Feng, J. Sun, W. Zhang, and D. Infield. Applying thermophysics for wind turbine drivetrain fault diagnosis using scada data. *IET Renewable Power Generation*, 10(5):661–668, 2016. doi:10.1049/iet-rpg.2015.0160. (cited on page 42)

[QGY06] Z. Qian, W. S. Gao, and Z. Yan. Application of case-based reasoning theory in fault diagnosis of power transformer. In *2006 IEEE 8th International Conference on Properties applications of Dielectric Materials*, pages 242–245, June 2006. doi:10.1109/ICPADM.2006.284162. (cited on page 71)

[RBM01] M. Redfern, Z. Bo, and D. Montjean. Detection of broken conductors using the positional protection technique. In *Power Engineering Society Summer Meeting, 2001*, volume 2, pages 1163–1168 vol.2, 2001. doi:10.1109/PESS.2001.970229. (cited on page 26)

[Red13] Red Hat Inc. JBoss Enterprise BRMS Platform 5 JBoss Rules 5 Reference Guide, 2013. (cited on pages 63, 92, 94)

[Rel12] Ieee guide for electric power distribution reliability indices - redline. *IEEE Std 1366-2012 (Revision of IEEE Std 1366-2003) - Redline*, pages 1–92, May 2012. (cited on page 2)

[RKD+11] S. E. Rudd, J. D. Kirkwood, E. M. Davidson, S. M. Strachan, V. M. Catterson, and S. D. J. McArthur. Intelligent monitoring of the health and performance of distribution automation. In *Intelligent System Application to Power Systems (ISAP), 2011 16th International Conference on*, pages 1–6, Sept 2011. doi:10.1109/ISAP.2011.6082220. (cited on page 45)

[RMJ10a] S. Rudd, S. Mcarthur, and M. Judd. A generic knowledge-based approach to the analysis of partial discharge data. *Dielectrics and Electrical Insulation, IEEE Transactions on*, 17(1):149–156, February 2010. `doi:10.1109/TDEI.2010.5412013`. (cited on pages ix, 64, 65)

[RMJ10b] S. Rudd, S. D. J. Mcarthur, and M. D. Judd. A generic knowledge-based approach to the analysis of partial discharge data. *IEEE Transactions on Dielectrics and Electrical Insulation*, 17(1):149–156, February 2010. `doi:10.1109/TDEI.2010.5412013`. (cited on page 6)

[RN10] S. Russell and P. Norvig. *Artificial intelligence - a modern approach.* Pearson Education, 2010. (cited on pages viii, 50, 51, 52, 54, 60)

[Rud10] S. E. Rudd. *Knowledge-Based Analysis of Partial Discharge Data*. PhD thesis, University of Strathclyde, 2010. (cited on pages 57, 61, 63)

[SAA+00] G. Schreiber, H. Akkermans, A. Anjewierdn, R. Hogg, N. Shadbolt, W. Velde, and B. Wielinga. *Knowledge engineering and management, the CommonKADS methodology.* The MIT Press, 2000. (cited on page 100)

[SBF11] R Studer, V. R. Benjamins, and D. Fensel. Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, 25(1):161–197, 2011. `doi:10.1016/S0169-023X(97)00056-6`. (cited on page 100)

[SBG02] E. Styvaktakis, M. Bollen, and I. Gu. Expert system for classification and analysis of power system events. *Power Delivery, IEEE Transactions on*, 17(2):423–428, Apr 2002. `doi:10.1109/61.997911`. (cited on pages 6, 63)

[Sch12] Schneider Electric Telemetry & Remote SCADA Solutions. Ethernet for SCADA Systems. Technical report, 2012. (cited on page 21)

[SdOF+08] R. H. Salim, K. R. C. de Oliveira, A. D. Filomena, M. Resener, and A. S. Bretas. Hybrid fault diagnosis scheme implementation for power distribution systems automation. *IEEE Transactions on Power Delivery*, 23(4):1846–1856, Oct 2008. doi:10.1109/TPWRD.2008.917919. (cited on page 3)

[SFK06] K. Stouffer, J. Falco, and K. Kent. Guide to Supervisory Control and Data Acquisition (SCADA) and Industrial Control Systems Security. Technical report, National Institute of Standards and Technology, 2006. (cited on page 22)

[SGR06] H. Salazar, R. Gallego, and R. Romero. Artificial neural networks and clustering techniques applied in the reconfiguration of distribution systems. *IEEE Transactions on Power Delivery*, 21(3):1735–1742, July 2006. doi:10.1109/TPWRD.2006.875854. (cited on page 55)

[SKK+96] J. Shinohara, I. Kozakai, M. Kunugi, J. Nagata, and H. Saito. Knowledge-based behavior interface: its application to power network restoration support system. *IEEE Transactions on Power Systems*, 11(1):383–389, Feb 1996. doi:10.1109/59.486122. (cited on page 6)

[SKS+00] E. Senger, W. Kaiser, J. Santos, P. Burt, and C. Malagodi. Broken conductors protection system using carrier communication. *Power Delivery, IEEE Transactions on*, 15(2):525–530, Apr 2000. doi:10.1109/61.852979. (cited on page 26)

[SLC+92] Y. Sun, C. C. Liu, R. D. Christie, J. Nordstrom, M. Hofmann, G. Stemler, and I. Thurein. Retex (relay testing expert): an expert system for analysis of relay testing data. *IEEE Transactions on Power Delivery*, 7(2):986–994, Apr 1992. doi:10.1109/61.127108. (cited on page 5)

[SLK+00] D. R. Sevcik, R. B. Lunsford, M. Kezunovic, Z. Galijasevic, S. Banu, and T. Popovic. Automated analysis of fault records and dissemination of event reports, 2000. (cited on page 44)

[SM11a] Abdelhay A. Sallam and Om P. Malik. *Main Concepts of Electric Distribution Systems*, pages 1–32. Wiley-IEEE Press,

2011. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5733139, doi:10.1002/9780470943854.ch1. (cited on page 11)

[SM11b] Abdelhay A. Sallam and Om P. Malik. *Protection of Electric Distribution Systems*, pages 163–235. Wiley-IEEE Press, 2011. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5733130, doi:10.1002/9780470943854.ch5. (cited on pages 29, 31, 31)

[SM11c] Abdelhay A. Sallam and Om P. Malik. *Scada Systems and Smart Grid Vision*, pages 469–493. Wiley-IEEE Press, 2011. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5733135, doi:10.1002/9780470943854.ch13. (cited on pages 1, 2, 2, 19)

[Sma14] Smart Grid Forum. Smart Grid Vision and Routemap. Technical report, ofgem, Department of Energy & Climate Change, 2014. (cited on page 1)

[SMB+10] J. Sykes, V. Madani, J. Burger, M. Adamiak, and W. Premerlani. Reliabilty of protection systems (what are the real concerns). In *Protective Relay Engineers, 2010 63rd Annual Conference for*, pages 1–16, March 2010. doi:10.1109/CPRE.2010.5469482. (cited on page 3)

[SMF01] M. Stanek, M. Morari, and K. Frohlich. Model-aided diagnosis: an inexpensive combination of model-based and case-based condition assessment. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 31(2):137–145, May 2001. doi:10.1109/5326.941838. (cited on page 71)

[Smi13] R. G. Smith. Knowledge-based systems concepts, techniques, examples, 2013. URL: http://www.reidgsmith.com/Knowledge-Based_Systems_-_Concepts_Techniques_Examples_08-May-1985.pdf. (cited on page 61)

[SP 04] SP Power Systems Ltd, ScottishPower Plc. NETWORK MANAGEMENT SYSTEMS FOR ACTIVE DISTRIBUTION NETWORKS - A FEASIBILITY STUDY. Technical report, 2004. (cited on page 21)

[SP 12a] SP Energy Networks. Assessment of overhead line performance during severe storms, 2012. URL: `http://www.spenergynetworks.co.uk/userfiles/file/201204_A2_6_EATL_Assmt_OH_line_perf_sev_storms.pdf`. (cited on page 77)

[SP 12b] SP Energy Netwroks. Application of overhead line switchgear and protection systems. Technical report, 2012. (cited on page 32)

[SP 13] SP Energy Networks. SP Long Term Development Statements for the years 2013/14 to 2017/18. Technical report, 2013. (cited on page 40)

[SP 15] SP Energy Networks. Innovation Funding Incentive Annual Report. Technical report, 2015. (cited on page 25)

[SRM+08] S. M. Strachan, S. Rudd, S. D. J. McArthur, M. D. Judd, S. Meijer, and E. Gulski. Knowledge-based diagnosis of partial discharges in power transformers. *IEEE Transactions on Dielectrics and Electrical Insulation*, 15(1):259–268, February 2008. `doi:10.1109/T-DEI.2008.4446759`. (cited on page 6)

[SRS+12] V. Spitsa, X. Ran, R. Salcedo, J. F. Martinez, R. E. Uosef, F. de Leon, D. Czarkowski, and Z. Zabar. On the transient behavior of large-scale distribution networks during automatic feeder reconfiguration. *IEEE Transactions on Smart Grid*, 3(2):887–896, June 2012. `doi:10.1109/TSG.2012.2186319`. (cited on page 3)

[SSLFF15] O. K. Siirto, A. Safdarian, M. Lehtonen, and M. Fotuhi-Firuzabad. Optimal distribution network automation considering earth fault events. *IEEE Transactions on Smart Grid*, 6(2):1010–1018, March 2015. `doi:10.1109/TSG.2014.2387471`. (cited on page 3)

[Sta09] Rudolph Frederick Stapelberg. *Handbook of reliability, availability, maintainability and safety in engineering design.* Springer, 2009. (cited on page 38)

[Str06] S. M. Strachan. *Decision Support System Development for Plant Condition Monitoring and Protection Design Reuse using Structured Methods of Knowledge Engineering and Knowledge Discovery.* PhD thesis, University of Strathclyde, 2006. (cited on page 57)

[SW10] C. L. Smallwood and J. Wennermark. Benefits of distribution automation. *IEEE Industry Applications Magazine*, 16(1):65–73, January 2010. `doi:10.1109/MIAS.2009.934970`. (cited on page 17)

[SWH94] G. Schreiber, B. Wielinga, and R. Hogg. CommonKADS: a comprehensive methodology for KBS development. *IEEE Expert*, pages 28–37, 1994. (cited on page 100)

[TCK+00] J. Tan, P. Crossley, D. Kirschen, J. Goody, and J. Downes. An expert system for the back-up protection of a transmission network. *Power Delivery, IEEE Transactions on*, 15(2):508–514, Apr 2000. `doi:10.1109/61.852976`. (cited on page 63)

[TG13] C. Tong and P. Guo. Data mining with improved apriori algorithm on wind generator alarm data. In *2013 25th Chinese Control and Decision Conference (CCDC)*, pages 1936–1941, May 2013. `doi:10.1109/CCDC.2013.6561250`. (cited on page 24)

[THL14] J. H. Teng, W. H. Huang, and S. W. Luan. Automatic and fast faulted line-section location method for distribution systems based on fault indicators. *IEEE Transactions on Power Systems*, 29(4):1653–1662, July 2014. `doi:10.1109/TPWRS.2013.2294338`. (cited on page 3)

[Tho03] J. Thomas. *Set theory.* Springer, 2003. (cited on page 132)

[TKV05] D. Thukaram, H. P. Khincha, and H. P. Vijaynarasimha. Artificial neural network and support vector machine approach for locating faults in radial distribution systems. *IEEE Transactions on Power Delivery*, 20(2):710–721, April 2005. `doi:10.1109/TPWRD.2005.844307`. (cited on pages viii, 55, 56)

[TLZ+14] Kang Taifeng, Ge Liang, Tan Zhihai, Zhao Fengqing, Huang Hongyuan, Peng Feijin, Cao Zhihui, and Ouyang Qijian. An alarm management framework for self-healing smart distribution system. In *2014 China International Conference on Electricity Distribution (CICED)*, pages 293–296, Sept 2014. doi:10.1109/CICED.2014.6991713. (cited on page 42)

[TM15] M. S. Thomas and J. D. McDonald. *Power System SCADA and Smart Grids*. CRC Press, 2015. (cited on page 22)

[TSK06] P.N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Always learning. Pearson Addison Wesley, 2006. URL: https://www-users.cs.umn.edu/~kumar/dmbook/ch8.pdf. (cited on page 57)

[TT15] H. T. T. Thein and K. M. M. Tun. Evaluation of differential evolution and k-means algorithms on medical diagnosis. In *Information Technology: Towards New Smart World (NSITNSW), 2015 5th National Symposium on*, pages 1–4, Feb 2015. doi:10.1109/NSITNSW.2015.7176408. (cited on page 59)

[VDMPVdH09] Laurens Van Der Maaten, Eric Postma, and Jaap Van den Herik. Dimensionality reduction: a comparative. *J Mach Learn Res*, 10:66–71, 2009. (cited on page 137)

[Ver16] H. K. Verma. FUNCTIONS OF SUPERVISORY CONTROL & DATA ACQUISITION (SCADA) SYSTEM, 2016. URL: http://profhkverma.info/sites/default/files/studentsData/scada/chapter2.pdf. (cited on page 23)

[VF98] Z. A. Vale and M. F. Ferandes. Better kbs for real-time applications in power system control centers: The experience of the sparse project. *Computers in Industry*, 37(2):97–111, Feb 1998. doi:10.1016/S0166-3615(98)00086-4. (cited on page 44)

[VFS13] A. Vukojevic, P. Frey, and M. Smith. Making a smart grid case: Fault circuit indicators. In *2013 IEEE Power Energy Society General Meeting*, pages 1–5, July 2013. doi:10.1109/PESMG.2013.6672436. (cited on page 2)

[Wil04] H. L. Willis. *Power Distribution Planning Reference Book.* CRC Press, 2004. (cited on page 13)

[WMS+17] X. Wang, S. McArthur, S. Strachan, J. Kirkwood, and B. Paisley. A data analytic approach to automatic fault diagnosis and prognosis for distribution automation. *IEEE Transactions on Smart Grid*, PP(99):1–1, 2017. doi:10.1109/TSG.2017.2707107. (cited on pages 86, 91, 115, 118, 135, 141)

[WSKM14] X. Wang, S. Strachan, J. Kirkwood, and S. McArthur. Automatic analysis of pole mounted auto-recloser data for fault prognosis to mitigate customer supply interruptions. In *Power Engineering Conference (UPEC), 2014 49th International Universities*, pages 1–6, Sept 2014. doi:10.1109/UPEC.2014.6934653. (cited on pages 35, 115)

[WSM+01] G. West, S. Strachan, A. Moyes, J. McDonald, B. Gwyn, and J. Farrell. Knowledge management and decision support for electrical power utilities. *Knowledge and Process Management*, 8(4):207–216, 2001. doi:10.1002/kpm.124. (cited on page 71)

[WSMK15] 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 *Intelligent System Application to Power Systems (ISAP), 2015 18th International Conference on*, pages 1–6, Sept 2015. doi:10.1109/ISAP.2015.7325519. (cited on page 36)

[WVKW16] T. Werho, V. Vittal, S. Kolluri, and S. M. Wong. A potential island formation identification scheme supported by pmu measurements. *IEEE Transactions on Power Systems*, 31(1):423–431, Jan 2016. doi:10.1109/TPWRS.2015.2402130. (cited on page 3)

[XC06] L. Xu and Mo-Yuen Chow. A classification approach for power distribution systems fault cause identification. *IEEE Transactions on Power Systems*, 21(1):53–60, Feb 2006. doi:10.1109/TPWRS.2005.861981. (cited on page 55)

[XK09] F. Xu and M. Kezunovic. Automated integration of substation ied data to generate event reports for protection engineers. In *North American Power Symposium (NAPS), 2009*, pages 1–6, Oct 2009. doi:10.1109/NAPS.2009.5484039. (cited on page 47)

[XLS+15] J. Xie, C. C. Liu, M. Sforna, M. Bilek, and R. Hamza. Intelligent physical security monitoring system for power substations. In *Intelligent System Application to Power Systems (ISAP), 2015 18th International Conference on*, pages 1–6, Sept 2015. doi:10.1109/ISAP.2015.7325524. (cited on page 17)

[Xu94] L. Xu. Case based reasoning. *Potentials, IEEE*, 13(5):10–13, Dec 1994. doi:10.1109/45.464654. (cited on pages 69, 70)

[Y. 13] Y. Paithankar and S. Bhide. *Fundamentals of power system protection*. PHI Learning Private Limited, 2013. (cited on pages xii, 26, 28)

[ZBWD05] Wei Zhao, Xiaomin Bai, Wenping Wang, and Jian Ding. A novel alarm processing and fault diagnosis expert system based on bnf rules. In *2005 IEEE/PES Transmission Distribution Conference Exposition: Asia and Pacific*, pages 1–6, 2005. doi:10.1109/TDC.2005.1546875. (cited on page 24)

[ZDGK10] Ce Zheng, Yimai Dong, O. Gonen, and M. Kezunovic. Data integration used in new applications and control center visualization tools. In *IEEE PES General Meeting*, pages 1–7, July 2010. doi:10.1109/PES.2010.5588140. (cited on page 46)

[ZL98] Y. Zhang and H. Liu. Application of plc and scada in auto-control systems for silo grain handling. In *Proceedings of the 7th International Working Conference on Stored-product Protection*, volume 2, pages 1919–1922, Oct 1998. (cited on page 24)