Building Knowledge for Substation-Based Decision Support Using Rough Sets

Ching-Lai Hor, *Member, IEEE*, Peter A. Crossley, *Member, IEEE*, and Simon J. Watson, *Member, IEEE*

***Abstract*—This paper describes a new technique based on rough sets to extract decision rules from large volumes of data captured by protection, control, and monitoring intelligent electronic devices. The methodology correctly identifies faults from large datasets and could be used to assist operators in their decision-making processes. Building knowledge for a fault diagnostic system is a time-consuming and costly process. The quality of a knowledge base can sometimes be hampered by a large number of superfluous decision-making rules that can lead to an unnecessarily large knowledge base system and inefficient or even detrimental rule maintenance. The methodology proposed cannot only induce decision rules efficiently but can also reduce the size of the knowledge base without causing loss of useful information. Results can be used by an expert system to generate supervisory automation and to support operators, for example, during an emergency situation. This methodology involves the generation of human–machine interface alarms. These can then be used for diagnosis of the type and cause of a fault event to give suggestions for network restoration and post-emergency repair. A Power Systems Computer Aided Design/Electromagnetic Transients including dc simulator has been used to investigate the effect of faults and switching actions on the protection and control equipment associated with a typical distribution network. The fundamental ideas of rough set theory are discussed, followed by a rule assessment method that is outlined using an illustrative example.**

***Index Terms*—Circuit breakers (CBs), discernibility, fault section estimation, relays, rough sets, rules discovery, voting algorithm.**

I. Introduction

THE amount of operational data captured within an electrical substation has increased significantly over recent years and human inspection and interpretation may no longer be feasible [1]. In many cases, operators often find themselves having only a vague idea of which parameters are important for their analysis. The main requirement for extracting concise and useful information is to determine the significant attributes of a data set by filtering out those attributes which are unimportant [2]. This paper proposes a novel method to extract knowledge from substation data acquired from relays and circuit breakers (CBs). The technique involved a process which learns and extracts knowledge from a set of events into a form of rules which are then validated using a voting algorithm before being verified by domain experts.

Manuscript received April 11, 2006. Paper no. TPWRD-00203-2006.

C.-L.Hor is with the Camborne School of Mines, School of Geography, Archaeology and Earth Resources, University of Exeter in Cornwall, Cornwall TR10 9EZ, U.K. (e-mail: C.L.Hor@exeter.ac.uk).

P. A. Crossley is with the Electrical Energy and Power Systems Group, School of Electrical and Electronic Engineering, University of Manchester, Manchester M13 9PL, U.K. (e-mail: p.crossley@manchester.ac.uk).

S. J. Watson is with the Center for Renewable Energy Systems Technology (CREST), Angela Marmont Renewable Energy Laboratory (AMREL), Loughborough University, Loughborough, Leicestershire LE11 3TU, U.K. (e-mail: s.j.watson@lboro.ac.uk).

Digital Object Identifier 10.1109/TPWRD.2006.886783

II. Rough Sets

A. Overview

Rough set theory is a soft computing technique for knowledge discovery that has been successfully applied to many areas of data analysis (e.g., medicine [3] and stock market analysis [4]). It consists of two main parts: dispensable attributes reduction and rules extraction. Two examples will be described to demonstrate how rough sets and a discernibility matrix are used to compute reducts. A reduct is a reduced set of attributes (e.g., voltages and currents) that essentially provides the same amount of information about a data set as a complete set of attributes. A relative discernibility matrix is then applied to this minimal attribute set to look for the core before any rules are extracted. A core is the set of relations occurring in every reduct (i.e., the set of all indispensable relations that characterize the equivalence relation).

The concept of rough sets is based on the idea of using upper and lower approximations of a set to deal with indiscernibility (see Section II-C for details). These two approximations provide crisp and rough descriptions of a data set. If a universe can be formed as a union of some elementary sets, it is called crisp; otherwise, it is rough. A lower approximation defines the collection of events in which the equivalence classes are fully contained in the set of events which is to be reduced to its essential attributes. The upper approximation, however, defines the collection of events in which the equivalence classes are at least partially contained in the set of events to be reduced. The approximations can also be divided into the positive (lower approximation), negative (complement of the upper approximation), and boundary (difference between the upper and lower approximation) regions.

In [5] and [6], the use of supervised and unsupervised rough classification, respectively, was proposed for handling large numbers of messages received during an emergency in order to reduce the quantity of data while maintaining useful and concise information. However, in the case of knowledge base construction for online intelligent switching, fault identification, and service restoration, detailed rules are required to cover every possible scenario which may occur in the substation. A classical information system may not cope well with high volumes of data, and the existing system may rely on particular data sources. If these data are not available or missing, this type of system may not perform accurately. Consequently, state estimation is required to calculate missing voltage and power

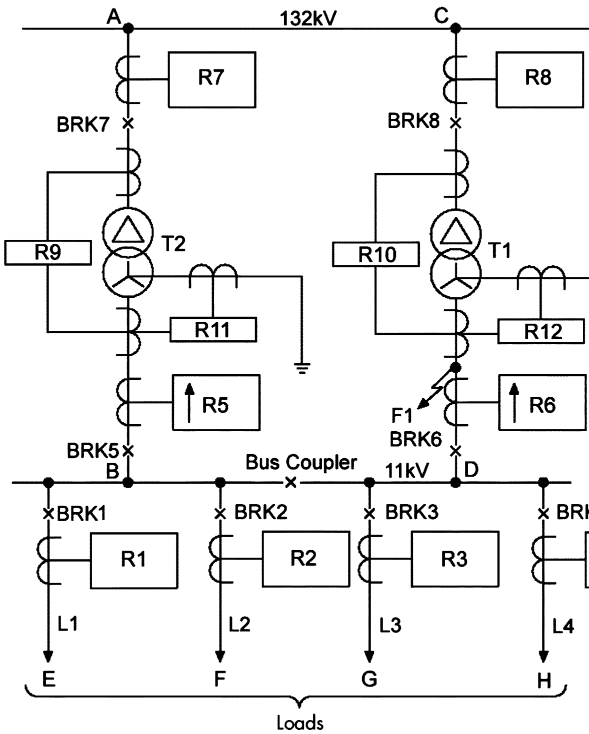

Fig. 1. The 132/11-kV substation model.

data across the entire network. To overcome these problems, we propose a new classification method based on rough sets for knowledge-base reduction and rule induction.

B. Decision System

A decision system acts upon a two-dimensional data matrix U. Each row in the universe matrix U corresponds to an event and each column represents an attribute. The decision system can be formulated as $\mathcal{D} = \{U, C \cup D\}$ where $\boldsymbol{U}$ represents a set of time events. $\boldsymbol{C}$ defines a set of condition attributes (i.e., observations) and $\boldsymbol{D}$ is a decision attribute that contains preclassified events. Any combination of the values for the decision attribute in $\mathbf{D}$ is represented by $\boldsymbol{d}$. Thus, $\mathbf{D} = \{d\}$. To illustrate the application of a decision system, a typical 132/11-kV substation model, as shown in Fig. 1, was developed using PSCAD/EMTDC [7]. The directional relays at R5 and R6 also include nondirectional time-graded earth-fault elements to protect the 11-kV busbar and provide backup for the 11-kV feeders [8]. Relays R1, R2, R3, and R4 all gave an identical pattern for the fault F1 and, therefore, these relays are regarded as one and labeled as "Rx" in which $x = \{1, 2, 3, 4\}$ (Table I). Vx and Ix represent the three-phase voltage and current. Similarly, breakers BRK6 and BRK8 are regarded as one protection zone labeled as "BZ3." H1 indicates that the current is flowing in the direction that would trigger one of the directional relays (i.e., R6).

d_1 indicates the state classifications. Normal $(\mathbf{N})$ indicates that all of the constraints and loads are satisfied (i.e., the voltages and currents are nominal (N)). Alert $(\mathbf{A})$ indicates at least one current is high (H) and the voltages are nominal, or the currents are nominal but at least one voltage is abnormal. Emergency $(\mathbf{E})$ indicates that at least two physical operating limits

TABLE I
DECISION SYSTEM

Rx		R5		R6		R7		R8		***d***
V_x	I_x	V_5	I_5	V_6	I_6	V_7	I_7	V_8	I_8	d_1d_2
N	N	N	N	N	N	N	N	N	N	**N0**
L	N	N	N	N	N	N	N	N	N	**A0**
L	N	L	N	L	N	N	N	N	N	**A0**
L	N	L	H	L	N	N	H	N	H	**E0**
L	N	L	H	L	H1	N	H	N	H	**E0**
L	L	L	H	L	H1	N	H	N	H	**E0**
L	L	L	H	L	H1	N	H	N	H	**E1**
L	N	L	H	L	H1	N	H	N	H	**E1**
L	N	L	H	L	L	N	H	L	N	**E1**
L	N	L	N	L	L	N	N	L	L	**A1**
N	N	L	N	L	L	N	N	L	L	**A1**
N	N	N	N	L	L	N	N	L	L	**S1**

are violated (e.g., under-(L) voltages and over-(H) currents). Safe $(\mathbf{S})$ is when those parts of the power system that are not isolated by the relays are operating normally, but one or more loads are not satisfied after a breaker has opened [9]. $d_2 = 1$ indicates that a breaker has opened and the associated line has been disconnected. Table I displays the voltage and current patterns captured by the relays in the event of fault F1. For brevity, only the change of state is presented. R9, R10, R11, and R12 are excluded from Table I since these unit protection relays do not contribute to this fault (F1) analysis. The time stamp in Table I is also not displayed. In addition, "BZ3" is omitted as it provides the same information as d_2.

C. Discernibility

"Discernibility" is the main theme of rough set analysis, defined as the ability to discern events from each other. It requires understanding how the characteristics of one event differ from another before the events can be classified. To achieve this, a discernibility matrix is used.

A discernibility matrix is a symmetric $n \times n$ matrix where n denotes the number of elementary sets [10]. All of the events in the rows and the columns are listed in the same order. In each entry of the matrix, the differences between the event corresponding to the row and the event corresponding to the column are compared and recorded. Naturally, the matrix will be symmetric due to the fact that the attribute, which differs in value for events a and b differs the other way around in value for events b and a [11]. Before defining the discernibility function, Table I must be converted into a discernibility matrix as shown in Table II. A discernibility function $f(B)$ is a Boolean function that expresses how an event (or a set of events) can be discerned from a certain subset of the full universe of events. A Boolean expression normally consists of Boolean variables and constants, linked by disjunction $(\bigvee)$ operators [11]. Given a decision system $\mathcal{D} = (U, B \cup \{d\})$, the discernibility function is

$$f_B^d(x_i) = \bigwedge \left\{ \bigvee \bar{m}_B^d(x_i, x_j) : 1 \leq j \leq i \leq n \right\} \qquad (1)$$

where $(\bigvee)$ and $(\bigwedge)$ are the disjunction and conjunction operators. $n = |U/\mathrm{IND}(B)|$, where $\mathrm{IND}(B)$ is the indiscernibility relation that partitions the objects U into a family of disjoint

TABLE II
DISCERNIBILITY MATRIX

	1	2	3	⋯	10	11	12
1	∅						
2	x	∅					
3	x,5,6	∅	∅				
4	x-8	5-8	5,7,8	⋯			
5	x-8	5-8	5-8	⋯			
6	x-8	x-8	x-8	⋯			
7	x-8B	x-8B	x-8B	⋯			
8	x-8B	5-8B	5-8B	⋯			
9	x-8B	5-8B	5-8B	⋯			
10	x,5,6,8,B	5,6,8,B	6,8,B	⋯	∅		
11	5,6,8,B	x,5,6,8,B	x,6,8,B	⋯	∅	∅	
12	6,8,B	x,6,8,B	x,5,6,8,B	⋯	x,5	5	∅

equivalence classes denoted as $U/\mathrm{IND}(B)$, which is indistinguishable from any other objects using only the available attributes in B. $\bigvee \bar{m}_B^d(x_i, x_j)$ is the disjunction taken over the set of Boolean variables $\bar{m}_B^d(x_i, x_j)$ corresponding to the variables $m_B^d(x_i, x_j)$ which is not equal to an empty set $\emptyset$ [12]. The decision relative discernibility function of B can be constructed to discern an event belonging to another class such as for an event class $x_k = (1 \le k \le n)$ over attributes B. This can be represented by the following function:

$$f(x_k, B) = \bigwedge \left\{ \bigvee \bar{m}_B^d(x_k, x_j) : 1 \le j \le n \right\}. \quad (2)$$

This function computes the minimal set of attributes B necessary to distinguish x_k from other event classes defined by B [12].

III. Rule Accuracy and Assessment

A decision rule can be denoted $\alpha \to \beta$, read as "if α, then β." The pattern α is called the rule's antecedence while the pattern β is called the rule's consequence. Three metrics as described below can be used to evaluate the quality of a given decision rule [13].

1) *Support*: The number of events that possesses both property α, then β.
2) *Accuracy*: A decision rule $\alpha \to \beta$ may only partially reveal the overall picture of the derived decision system. Given pattern α, the probability of the conclusion β can be assessed by measuring how trustworthy the rule is in drawing the conclusion β on the basis of evidence α

$$\text{Accuracy}(\alpha \to \beta) = \frac{\text{support}(\alpha \cdot \beta)}{\text{support}(\alpha)}. \quad (3)$$

3) *Coverage*: The strength of the rule relies upon the large support basis that describes the number of events, which support each rule. The quantity coverage $(\alpha \to \beta)$ is required in order to measure how well the evidence α describes the decision class. It can be defined via β

$$\text{Coverage}(\alpha \to \beta) = \frac{\text{support}(\alpha \cdot \beta)}{\text{support}(\beta)}. \quad (4)$$

IV. Voting Algorithm

There are various ways of classifying events using rule sets, and a voting algorithm can be used to resolve the conflicts and rank the predicted outcomes. This works reasonably well for rule-based classification.

Let RUL denote an unordered set of decision rules. The voting process is a way of employing RUL to assign a certainty factor to each decision class for each event. The concept of the voting algorithm can be divided into three parts [13].

1) The set of rules RUL searches for applicable rules $\mathrm{RUL}(x)$ that match the attributes of event x (i.e., rules that fire) in which $\mathrm{RUL}(x) \subseteq \mathrm{RUL}$.
2) If no rule is found (i.e., $\mathrm{RUL}(x) = \emptyset$), no classification will be made. The most frequently occurring decision is chosen. If more than one rule fires, this means that more than one possible outcome exists.
3) The voting process is performed in three stages:
   a) *Casting the votes*: Let a rule $r \in \mathrm{RUL}(x)$ cast as many votes, votes (r) in favor of its outcomes associated with the support counts as given

$$\text{votes}(r) = |||\alpha \cap \beta|||. \quad (5)$$

   b) *Compute a normalization factor, norm(x)*. The normalization factor is computed as the total number of votes cast and only serves as a scaling factor

$$\text{norm}(x) = \sum_{r \in \mathrm{RUL}(x)} \text{votes}(r_i). \quad (6)$$

   c) *Certainty Coefficient*: The votes from all of decision rules β are accumulated before they are divided by the normalization factor norm (x) to yield a numerical certainty coefficient. Certainty (x, β) for each decision class is given

$$\text{Certainty}(x, \beta) = \left(\frac{\text{votes}(\beta)}{\text{norm}(x)} \right) \quad (7)$$

   in which the $\text{votes}(\beta) = \sum \{\text{votes}(r)\}$ and $r \in \mathrm{RUL}(x) \wedge r \equiv (\alpha \to \beta)$. The certainty coefficient decides which rules will be the best fit for the unknown event.

V. Example I

This example considers a fault scenario on the 11-kV transformer T1 feeder given in Fig. 1. The data were generated from PSCAD/EMTDC simulator software [7], [8]. The fault (F1) both results in the operation of the directional relay R6, the tripping of BRK6 and BRK8, and the isolation of the T1. The decision system in Table I is transformed into a discernibility matrix shown in Table II, where $\{x-8\} = \{x, 5, 6, 7, 8\}$, $\{x-8B\} = \{x, 5, 6, 7, 8, BZ3\}$, $\{5-8\} = \{5, 6, 7, 8\}$, $\{5-8B\} = \{5, 6, 7, 8, BZ3\}$.

Based on the discernibility functions derived from each column in Table II using (1), the final discernibility function $f(\mathcal{D}) = Rx \cdot R5 \cdot BZ3$, where "$\cdot$" refers to the conjunction operator $(\bigwedge)$. As $Rx = \{R1, R2, R3, R4\}$ and

TABLE III
REDUCT TABLE

Rule	Rx		R5		BRK	d	Support
No.	V_x	I_x	V_5	I_5	BZ3	d_1d_2	count
1	N	N	N	N	0	N0	1
2	L	N	N	N	0	A0	1
3	L	N	L	N	0	A0	1
4	L	N	L	H	0	E0	2
5	L	L	L	H	0	E0	1
6	L	L	L	H	1	E1	1
7	L	N	L	H	1	E1	2
8	L	N	L	N	1	A1	1
9	N	N	L	N	1	A1	1
10	N	N	N	N	1	S1	1

TABLE IV
QUALITY OF RULE MEASURE

Rule	Acc	LCov	RCov	LLH	RLH	LSP	RSP
1	1.0	0.08	1.00	3	1	1	1
2	1.0	0.08	0.50	3	1	1	1
3	1.0	0.08	0.50	3	1	1	1
4	1.0	0.17	0.67	3	1	2	2
5	1.0	0.08	0.33	3	1	1	1
6	1.0	0.08	0.33	3	1	1	1
7	1.0	0.17	0.67	3	1	2	2
8	1.0	0.08	0.50	3	1	1	1
9	1.0	0.08	0.50	3	1	1	1
10	1.0	0.08	1.00	3	1	1	1

Acc:Accuracy, LCov:LHS Coverage, RCov:RHS Coverage, LLH:LHS Length, RLH:RHS Length, LSP:LHS Support, RSP:RHS Support.

$BZ3 = \{BRK6, BRK8\}$, a total of eight reducts are generated. Depending on data availability, any of these reducts can be used to classify the events. In other words, if there are some missing data (e.g., R1 and R4 are not available), we can use the data from R2 or R3 and R5 and BRK6 or BRK8. The reduct set is given in Table III.

A. Quality of Rule Measure

The quality of rules from Table III can be assessed based on the metrics: right-hand side (RHS) and left-hand side (LHS) support, accuracy coverage, and length shown in Table IV. The LHS support signifies the number of events in the data set. The RHS support signifies the number of events in the data set that match the "if" part of the decision rule and have the decision value of the "then" part (consequent). For an inconsistent rule, the "then" part shall consist of several decisions. Accuracy and coverage are computed from the support counts using (3) and (4). Since there is no inconsistency in the decision system, the accuracy of rules is equal to 1.0. Length indicates the number of attributes in the LHS or RHS; LHS length $= 3(Rx, R5, BZ3)$ and RHS length $= 1$.

B. Relative Discernibility Functions

Table III may include some unnecessary values of the condition attributes. To condense the rules, the relative reduct and core are computed using the relative discernibility function in (2). These are based on the relative discernibility matrix constructed for the subspace $\{Rx, R5, BZ3\}$ (Table V).

In Table V, let $Rx = \{V_x, I_x\}$ and $R5 = \{V_5, I_5\}$. Voltage and current attributes in each relay are considered separately rather than treating them as one unit. In each column of Table V, the relative discernibility functions are computed. For example, to construct $f(1,B)$, where $B \subseteq A$, all sets of attributes from column 1 are summed using the absorption law, similar to $f(2,B)$ using all sets of attributes from column 2, etc. Here, "$\cdot$" refers to the conjunction operator $(\bigwedge)$ and "$+$ " refers to the disjunction operator $(\bigvee)$. The result for $f(2,B)$ indicates that there are two rules necessary to classify the abnormal state. The first rule requires the attributes $\{V_x, V_5, I_5\}$, whereas the second rule requires the attributes $\{V_x, V_5, BZ3\}$. The relative discernibility functions are converted into 16 decision rules. However, of these rules, two are actually redundant (i.e., 4 and $5^{(1)}$ are identical as are $6^{(1)}$ and 7). We have discarded $5^{(1)}$ and $6^{(1)}$, leaving only 14 applicable rules as listed in Table VI.

For simplicity, the rule numbers in Table VI are renamed Rule 1 – 13 (omitting the first rule which represents the normal operation and is not of interest for fault classification and event ex-

TABLE V
RELATIVE DISCERNIBILITY MATRIX

	1	2	$\cdots$	10
1	$\emptyset$	V_x	$\cdots$	BZ3
2	V_x	$\emptyset$	$\cdots$	V_x, BZ3
3	V_x, V_5	$\emptyset$	$\cdots$	V_x, V_5, BZ3
4	V_x, R_5	R_5	$\cdots$	V_x, R_5, BZ3
5	R_x, R_5	I_x, R_5	$\cdots$	R_x, R_5, BZ3
6	R_x, R_5, BZ3	I_x, R_5, BZ3	$\cdots$	R_x, R_5
7	V_x, R_5, BZ3	R_5, BZ3	$\cdots$	V_x, R_5
8	V_x, V_5, BZ3	V_5, BZ3	$\cdots$	V_x, V_5
9	V_5, BZ3	V_x, V_5, BZ3	$\cdots$	V_5
10	BZ3	V_x, BZ3	$\cdots$	$\emptyset$

$$
\begin{aligned}
f(1,B) &= V_x \cdot BZ3 \\
f(2,B) &= V_x \cdot (V_5 + I_5) \cdot (V_5 + BZ3) \\
&= (V_x \cdot V_5 \cdot I_5) + (V_x \cdot V_5 \cdot Bz3) \\
f(3,B) &= I_5 \cdot BZ3 \cdot (V_x + V_5) \\
&= (V_x \cdot I_5 \cdot BZ3) + (V_5 \cdot I_5 \cdot BZ3) \\
f(4,B) &= I_5 \cdot BZ3 \\
f(5,B) &= (I_x + I_5) \cdot BZ3 = (I_x \cdot BZ3) + (I_5 \cdot BZ3) \\
f(6,B) &= (I_x + I_5) \cdot BZ3 = (I_x \cdot BZ3) + (I_5 \cdot BZ3) \\
f(7,B) &= I_5 \cdot BZ3 \\
f(8,B) &= I_5 \cdot BZ3 \cdot (V_x + V_5) \\
&= (I_5 \cdot BZ3 \cdot V_x) + (I_5 \cdot BZ3 \cdot V_5) \\
f(9,B) &= V_5 \cdot (V_x + BZ3) \cdot (V_x + I_5) \\
&= (V_5 \cdot V_x \cdot I_5) + (V_5 \cdot V_x \cdot BZ3) \\
f(10,B) &= V_5 \cdot BZ3
\end{aligned}
$$

TABLE VI
CORE TABLE

Rule	Rx		R5		BRK	d	Sup.1	Sup.2
No.	V_x	I_x	V_5	I_5	BZ3	d_1d_2	Index	Index
1	N	•	•	•	0	N 0	1	1
2	L	•	N	N	•	A 0	1	1
$2^{(1)}$	L	•	N	•	0	A 0	1	1
3	L	•	•	N	0	A 0	2	2
$3^{(1)}$	•	•	L	N	0	A 0	1	1
4	•	•	•	H	0	E 0	3	5
5	•	L	•	•	0	E 0	1	2
6	•	L	•	•	1	E 1	1	3
7	•	•	•	H	1	E 1	3	8
8	L	•	•	N	1	A 1	1	2
$8^{(1)}$	•	•	L	N	1	A 1	2	3
9	N	•	L	N	•	A 1	1	1
$9^{(1)}$	N	•	L	•	1	A 1	1	1
10	•	•	N	•	1	S 1	1	1

Sup.1 Index: support count index 1 based on the number of events given in Table I.
Sup.2 Index: support count index 2 based on a more complete data set using a 3-phase current and voltage.

traction) and are categorized into five different classes according to their outcomes.

1) **ABNORMAL A0**

$$\begin{array}{llllll} \textbf{Rule 1}: & V_x(L), & V_5(N), & I_5(N) & Z1 \\ \textbf{Rule 2}: & V_x(L), & V_5(N), & BZ3(0) & Z1 \\ \textbf{Rule 3}: & V_x(L), & I_5(N), & BZ3(0) & Z1 \\ \textbf{Rule 4}: & V_5(L), & I_5(N), & BZ3(0) & Z25 \end{array}$$

The system behaves abnormally and is at high alert. Zone $Z1$ and $Z2$ both experience voltage sags. Note: the substation in Fig. 1 can be divided into four main protection zones. Zone 1 represents the protection zones of R1, R2, R3, and R4. Zone 2 includes the protection zones of R5, R7, R9, and R11. Zone 3 covers the zones of R6, R8, R10, and R12. Zone 4 is the busbar protection zone which is not considered in this scenario. Protection Zone 25 indicates that the regional Zone 2 is supervised by the relay R5.

2) **ABNORMAL A1**

$$\begin{array}{llllll} \textbf{Rule 5}: & V_x(L), & I_5(N), & BZ3(1) & Z1\&Z3 \\ \textbf{Rule 6}: & V_5(L), & I_5(N), & BZ3(1) & Z25\&Z3 \\ \textbf{\textit{Rule}}\ 7: & V_x(N), & V_5(L), & I_5(N) & Z25 \\ \textbf{Rule 8}: & V_x(N), & V_5(L), & BZ3(1) & Z25\&Z3. \end{array}$$

The system is recovering. Protection at Zone 3 has responded. The situation is under control but not safe.

3) **EMERGENCY E0**

$$\begin{array}{lllll} \textbf{Rule 9}: & I_5(H), & BZ3(0) & Z25 \\ \textbf{\textit{Rule}}\ \mathbf{10}: & I_x(L), & BZ3(0) & Z1 \end{array}$$

The system is unstable and urgent action is required. Protection has not yet responded.

4) **EMERGENCY E1**

$$\begin{array}{lllll} \textbf{\textit{Rule}}\ \mathbf{11}: & I_x(L), & BZ3(1) & Z1\&Z3 \\ \textbf{\textit{Rule}}\ \mathbf{12}: & I_5(L), & BZ3(1) & Z25\&Z3 \end{array}$$

The system is still unstable. Protection at Zone 3 has responded. The fault is isolated to Zone 3.

5) **SAFE S1**

$$\textbf{Rule 13}: \quad V_5(N), \quad BZ3(1) \quad Z3$$

The system is within the safe margin. A fault analysis report is generated that identifies the fault type and the affected region. The condition of the protection is evaluated. Restoration procedure and maintenance records are generated accordingly.

Rules 7, 10, 11, and 12 (in italics) may have to be modified as they do not clearly justify the status alarm. If the extracted rule does not comply with the state classification (normal, abnormal, emergency and safe) set earlier, it does not mean that the rules extraction is inaccurate, simply because the data set does not contain adequate information to classify the events. This limited knowledge of the data set can be noticed by comparing the unmatched rules with the original table.

The rules generated for each scenario are stored in the knowledge base system. The inference engine then uses a lookup table to retrieve the mapping between the input values and the rule's consequence(s) for each scenario. This means that if the fault symptom matches the list of rules given (the antecedences of the extracted rules), a fault in Zone Z36 (Zone 3 supervised by the relay R6) is concluded. Different sets of decisions could also be fired based on the rule's consequence(s). A lookup table can be thought of as a matrix, which has as many columns as there are inputs, and as many rows as outcomes. The inference engine thus takes a set of inputs and matches this input pattern with the patterns in the matrix, stored in rows. The best match determines the outcome as the proper answer. The matching process could sum the matching bits in a row, or carry out a multiplication of the input as a column vector with the matrix. The largest element in the product column selects the outcome. In a time-independent system, where the outcome depends only on the instantaneous state of the inputs, the case structure can be very conveniently expressed in this form of matrix. The advantage of using this method is that the rules can be induced easily. This saves significant time and cost when developing a knowledge base. The example shows that the approach is capable of inducing the decision rules from a substation data base, even though the data may not be complete.

C. Voting Results

Table VII illustrates the results computed by the voting algorithm. Assuming that only the rules presented with $V_5 = L$ and $I_5 = N$ were fired, the voting algorithm based on the support count index 1 concluded an ABNORMAL decision (combining the result of $A0 = 4/9$ and $A1 = 5/9$). The support count for the case $V_5 = L$ or $I_5 = N$ with outcome A0 is 4, whereas the total support count for the case $V_5 = L$ or $I_5 = N$ regardless of any outcome is 9. The same procedure applies to A1 in which the support count for the case $V_5 = L$ or $I_5 = N$ with outcome A1 is 5. With this set of rules, the most likely decision is an ABNORMAL state. The support count, however, shows a marginal preference for abnormal state A1. Let us now assume that only rules presented with $V_x = L$ and $V_5 = L$ and $I_5 = H$ fire. We have accumulated the casted votes for all rules that fire and divided them by the number of support count for all rules that fire (i.e., 16). The voting algorithm indicates that an abnormal state is the likely decision instead of the emergency state due to its higher support count in the given set of rules. This may not be agreed on by all experts. The reason for this conflict is caused by the inadequate information in the small dataset in Table I. As a result, the rule coverage is limited particularly during the emergency period. To support our conclusion, we apply the support count index 2 based on a more complete data set that contains three-phase currents and a three-phase voltage. The same procedure is repeated and this time, the emergency state is chosen as seen in Table VII with the certainty coefficients computed for each decision class. The suggestion from the voting result should always be left in the final analysis to domain experts to decide the necessary action to be taken.

TABLE VII
ACCUMULATING THE CASTED VOTES FOR ALL RULES THAT FIRE

Index	Certainty	Fraction	Decimal
1	certainty$(x, (d_1 d_2 = \mathrm{A0}))$	$5/16$	0.31
	certainty$(x, (d_1 d_2 = \mathrm{E0}))$	$3/16$	0.19
	certainty$(x, (d_1 d_2 = \mathrm{E1}))$	$3/16$	0.19
	certainty$(x, (d_1 d_2 = \mathrm{A1}))$	$5/16$	0.31
2	certainty$(x, (d_1 d_2 = \mathrm{A0}))$	$5/25$	0.20
	certainty$(x, (d_1 d_2 = \mathrm{E0}))$	$5/25$	0.20
	certainty$(x, (d_1 d_2 = \mathrm{E1}))$	$8/25$	0.32
	certainty$(x, (d_1 d_2 = \mathrm{A1}))$	$7/25$	0.28

D. Classifier Performance

The set of rules derived from the reducts must be assessed on its classification performance, readability, and usefulness before it can be used effectively for online diagnosis. Usually, a domain expert shall be the one who evaluates the usefulness of the rules because of his or her knowledge about the power system and experience from operating and monitoring the process. When classifying a new and unseen event using a set of rules, we generally expect to see a return value for each event from the classification algorithm. If rules are matched and able to classify all events, that decision is definitely chosen. However, in those cases where the rules are not able to classify all events, particularly when more than one classification is possible, the algorithm will then have to make an educated guess. A good guess may indeed correctly classify some of the undefined events. However, a wrong guess could result in wrong classification. In power systems, event classification is crucial. Wrong classification may lead to a dangerous situation in the worst case. Therefore, if the rules are not able to classify all events, the operators should be informed. This is because the operators would stand a better chance and are more qualified to make an educated guess than a classification algorithm.

For assessing the classifier performance, the dataset is divided into a training set and a test set. The training set is a set of examples used for learning that is used to fit the parameters, whereas the testset is a set of examples used only to assess the performance of a classifier. Rules are mined from a selection of events in each training set using rough sets. They are then used to classify the events in the testset. If the rules cannot classify the events in the testset satisfactorily, the rules must be notified to the user and refined to suit the real application. This method can be carried out for two purposes. First, the rule set can be viewed as a classifier, used for the purpose of classifying only. Second, the computed reducts and the generated rules can be used by domain experts to learn more about the data. The last approach often requires a small set of rules for the human expert to examine; thus, rule filtering can be carried out [14].

The original simulation data is randomly divided into three different training sets and test sets, respectively, with a partition of 90%, 70%, and 50% of the data for training and 10%, 30%, and 50% for testing. The procedure is repeated four times for four random splits of the data. This means that four different test sets were generated in each case and each of these was tested on every split of the training sets for a total of four runs. The splits are used to avoid results based on rules that were generated for a particular selection of events [15]. This makes the results more reliable and independent of one particular selection of events.

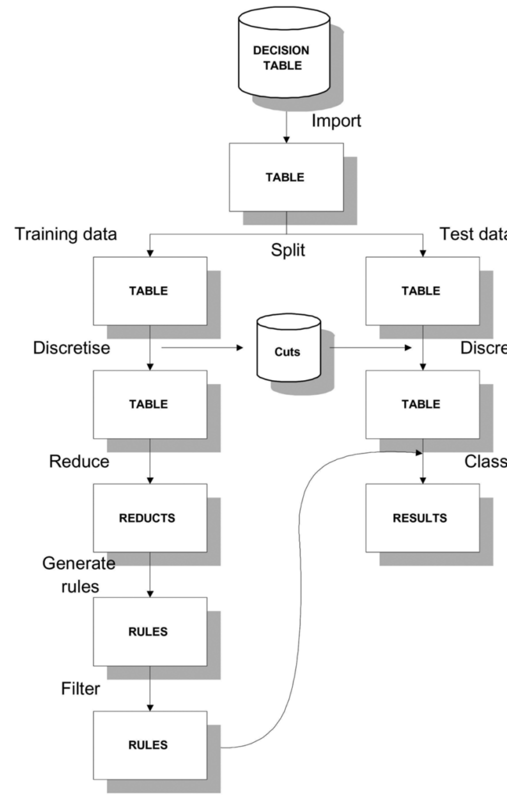

Fig. 2. Analysis steps for assessing rules performance.

TABLE VIII
CLASSIFIER RESULT USING THE 90% TRAINING SET AND 10% TESTSET

Training Set (90%)	Test Sets (10%) Split 1	Split 2	Split 3	Split 4	Mean Accuracy
1	1.000	1.000	1.000	1.000	1.000
2	1.000	1.000	1.000	1.000	1.000
3	1.000	1.000	1.000	1.000	1.000
4	1.000	1.000	1.000	1.000	1.000
	Measure of Accuracy				**1.000**

TABLE IX
CLASSIFIER RESULT USING THE 70% TRAINING SET AND 30% TESTSET

Training Set (70%)	Test Sets (30%) Split 1	Split 2	Split 3	Split 4	Mean Accuracy
1	1.000	1.000	1.000	1.000	1.000
2	1.000	1.000	1.000	1.000	1.000
3	1.000	1.000	1.000	1.000	1.000
4	1.000	1.000	1.000	1.000	1.000
	Measure of Accuracy				**1.000**

TABLE X
CLASSIFIER RESULT USING THE 50% TRAINING SET AND 50% TESTSET

Training Set (50%)	Test Sets (50%) Split 1	Split 2	Split 3	Split 4	Mean Accuracy
1	0.933	1.000	0.967	1.000	0.975
2	0.900	0.733	0.800	0.833	0.817
3	1.000	1.000	0.967	1.000	0.992
4	1.000	1.000	1.000	1.000	1.000
	Measure of Accuracy				**0.946**

However, sometimes the events are not equally distributed over the given decision classes when the dataset is split. One decision class may dominate over the other decision classes. Due to a lack of events in this analysis, to overcome this problem, the events are duplicated to ensure the same equal number of events are

TABLE XI
LIST OF VOLTAGE AND CURRENT PATTERNS WITH ESTIMATED PROTECTION ZONES FOR VARIOUS FAULT SCENARIOS

R1		R2		R3		R4		R5		R6		R7		R8		R9	R10	R11	R12	ZONE
V_1	I_1	V_2	I_2	V_3	I_3	V_4	I_4	V_5	I_5	V_6	I_6	V_7	I_7	V_8	I_8	I_9	I_{10}	I_{11}	I_{12}	
L	H	L	L	L	L	L	L	L	H	L	H	N	H	N	H	L	L	L	L	Z11
L	L	L	L	L	L	L	L	L	H1	L	H	N	H	N	H	L	L	L	L	Z25
L	L	L	L	L	L	L	L	L	H	L	H1	N	H	N	H	L	L	L	L	Z36
L	L	L	L	L	L	L	L	L	L	L	L	L	H	L	L	L	L	L	L	Z27
L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	H	L	L	L	L	Z38
L	L	L	L	L	L	L	L	L	H1	L	H	L	H	L	H	H	L	L	L	Z2T
L	L	L	L	L	L	L	L	L	H1	L	H	L	H	L	H	L	L	H	L	Z2T
L	L	L	L	L	L	L	L	L	H1	L	H	L	H	L	H	H	L	H	L	Z2T
L	L	L	L	L	L	L	L	L	H	L	H1	L	H	L	H	L	H	L	L	Z3T
L	L	L	L	L	L	L	L	L	H	L	H1	L	H	L	H	L	L	L	H	Z3T
L	L	L	L	L	L	L	L	L	H	L	H1	L	H	L	H	L	H	L	H	Z3T

TABLE XII
LIST OF SWITCHING ACTIONS WITH ESTIMATED PROTECTION ZONES FOR VARIOUS FAULT SCENARIOS

R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	B1	B2	B3	B4	BZ2	BZ3	ZONE
10	01	01	01	01	01	01	01	01	01	01	01	10	01	01	01	01	01	Z11
01	10	01	01	01	01	01	01	01	01	01	01	01	10	01	01	01	01	Z12
01	01	10	01	01	01	01	01	01	01	01	01	01	01	10	01	01	01	Z13
01	01	01	10	01	01	01	01	01	01	01	01	01	01	01	10	01	01	Z14
01	01	01	01	10	01	01	01	01	01	01	01	01	01	01	01	10	01	Z25
01	01	01	01	01	10	01	01	01	01	01	01	01	01	01	01	01	10	Z36
01	01	01	01	01	01	10	01	01	01	01	01	01	01	01	01	10	01	Z27
01	01	01	01	01	01	01	10	01	01	01	01	01	01	01	01	01	10	Z38
01	01	01	01	01	01	01	01	10	01	10	01	01	01	01	01	10	01	Z2T
01	01	01	01	01	01	01	01	10	01	01	01	01	01	01	01	10	01	Z2T
01	01	01	01	01	01	01	01	01	01	10	01	01	01	01	01	10	01	Z2T
01	01	01	01	01	01	01	01	01	10	01	10	01	01	01	01	01	10	Z3T
01	01	01	01	01	01	01	01	01	10	01	01	01	01	01	01	01	10	Z3T
01	01	01	01	01	01	01	01	01	01	01	10	01	01	01	01	01	10	Z3T

distributed over the decision classes in order to make the voting process more capable of detecting these events, thus providing better classification. Fig. 2 shows the complete process to determine how the rules can be induced and then classified to assess their performance.

Tables VIII and IX show that we have achieved 100% in the accuracy of classification for the 10% and 30% test set. Table X shows that when 50% of the data are used, the accuracy only dropped to 94.6%. The results revealed that the extracted rules have a high successful classification rate.

VI. EXAMPLE II

Tables XI and XII laid out a simple example containing a list of voltage and current patterns as well as the switching actions caused by the protection system(s) subject to various faults at different locations in the substation (Fig. 1). $Bx =$ the breaker x in which $x = \{1, 2, 3, 4\}$, such as BZ3, BRK5, and BRK7 being regarded as one and labeled "BZ2." The auxiliary contacts are used to determine the condition of a breaker and relay. "01" indicates that the contact of the breaker/relay is closed. "10" indicates that the breaker/relay is open/tripped, "00" indicates failure of the breaker/relay and "11" indicates an undefined breaker/relay state. This information about the auxiliary contacts is particularly useful when the protection system has failed/malfunctioned. $Ix = \{I_1, I_2, I_3, I_4\}$ since all of the load currents have similar patterns.

Going through the same procedure described in Section II and combining the information from Tables XII and XIII, six concise decision rules obtained can be interpreted.

TABLE XIII
RULES GENERATED FOR VARIOUS FAULT SCENARIOS IN THE SUBSTATION

I_x	I_5	V_7	I_7	ZONE
H	•	•	•	Z1x
•	H1	N	•	Z25
L	H	N	•	Z36
•	L	•	H	Z27
•	•	•	L	Z38
•	H1	L	•	Z2T
•	H	L	•	Z3T

RULE 1: IF $I_x = H$, $Rx = 10$ and $Bx = 10$, then the fault section lies within Zone 1x, in which $x = \{1, 2, 3, 4\}$.
RULE 2: IF $I_5 = H1$, $V_7 = N$, $R5 = 10$ and $BZ2 = 10$, then the fault section lies within Zone 25.
RULE 3: IF $I_x = L$, $I_5 = H$, $V_7 = N$, $R6 = 10$, and $BZ3 = 10$, then the fault section lies within Zone 36.
RULE 4: IF $I_7 = L$, $R8 = 10$, and $BZ3 = 10$, then the fault section lies within Zone 38.
RULE 5: IF $I_5 = H1$, $V_7 = L$, and $R9 = 10$ and/or $R11 = 10$ and $BZ2 = 10$, then the fault section lies within Zone 2T. Zone 2T is the region within the Zone 2 that is supervised by transformer unit protection.
RULE 6: IF $I_5 = H$, $V_7 = L$ and $R8 = 10$ and/or $R12 = 10$ and $BZ3 = 10$, then the fault section lies within Zone 3T.

The example given is small and incomplete. Therefore, some of these extracted rules may appear oversimplified. This is likely to occur when the dataset does not contain adequate information for knowledge extraction. The solution is either to acquire a more complete data set (which will not be a problem with the large quantity of data modern relays/IEDs can generate) or some of the rules should be refined by experts to improve the coverage. The results also look predictable for a small substation as in Fig. 1. However, when considering a larger substation or a complex power network with a large number of protection system(s), extracting rules manually may be time-consuming and require significant resources. As such, the method described in this paper will be useful to power utilities for exploiting substation rules. It could also help to reduce the size of conventional rule-base systems by eliminating superfluous information that may exist in the knowledge base. The rules produced are generally concise. Relying on the switching actions for fault section estimation might not always be adequate when considering relay failures and the complexity of a power network. Therefore, we believe that voltage and current components should also be considered in a fault section estimation procedure.

VII. Conclusion

This paper suggests the use of a novel, structured method to process and extract implicit knowledge from operational data derived from relays and circuit breakers (CBs). The technique has been applied to identify underlying data relationships and simplified logic-based rules that can be used to identify or classify fault section and abnormal events. The theoretical approach taken is simple but robust and the resulting method has shown promise for eventual application in the power system engineering domain. The methodology is more attractive than some other techniques such as the Bayesian approach because no assumption about the independence of the attributes is necessary nor is any background knowledge about the data required [16]. A set of training data of reasonable quality is needed. Though decision trees have been used successfully in ID3 and C4.5, compared to rule sets generated by rough set theory, it remains questionable whether decision trees can be described as knowledge, regardless of how well they function [17]. Their performance can also be affected by the presence of missing values in the test data set. This is less likely in the case of rough set theory.

Rules extraction and subsequent classification can be performed without the presence of an expert even though experts may still have to perform the final check before these rules are used in the real-time application. The technique simplifies rule generation (knowledge acquisition) and reduces the time and resources required to develop a rule-based diagnostic system. The extracted knowledge is a set of propositional rules which can be said to have syntactic and semantic simplicity for a human. Two examples have been given to show how knowledge can be deduced from datasets and from these simplified examples, the results show promise for practical applications.

References

[1] M. Eby, "Don't let data overload stop you," *Transm. Distrib. World*, May 1999.

[2] X. Hu, "Knowledge discovery in databases: An attribute-oriented rough set approach," Ph.D. dissertation, Dept. Comput. Sci., Univ. Regina, Regina, SK, Canada, 1995.

[3] A. Øhrn and T. Rowland, "Rough sets: A knowledge discovery technique for multifactorial medical outcomes," *Amer. J. Phys. Med. Rehab.*, vol. 79, no. 1, pp. 100–108, 2000.

[4] Y. Wang, "Mining stock price using fuzzy rough set system," *Expert Syst. Appl.*, vol. 24, no. 1, pp. 13–23, Jan. 2003.

[5] C. Hor and P. Crossley, "Extracting knowledge from substations for decision support," *IEEE Trans. Power Del.*, vol. 20, no. 2, pt. 2, Apr. 2005.

[6] ——, "Unsupervised event extraction within substations using rough classification," *IEEE Trans. Power Del.*, vol. 21, no. 4, pp. 1809–1816, Oct. 2006.

[7] C. Hor, A. Shafiu, P. Crossley, and F. Dunand, "Modelling a substation in a distribution network: Real time data generation for knowledge extraction," presented at the IEEE Power Eng. Soc. Summer Meeting, Chicago, IL, Jul. 21–25, 2002.

[8] C. Hor, K. Kangvansaichol, P. Crossley, and A. Shafiu, "Relay modelling for protection studies," presented at the IEEE Bologna Power Tech Conf., Bologna, Italy, Jun. 23–26, 2003.

[9] G. Gross, A. Bose, C. DeMarco, M. Pai, J. Thorp, and P. Varaiya, "Real time security monitoring and control of power systems." 1999, Tech. Rep., Consort. Elect. Rel. Technol. Solutions (CERTS) Grid of the Future White Paper.

[10] A. Skowron and C. Rauszer, "The discernibility matrices and functions in information systems," in *Intelligent Decision Support—Handbook of Applications and Advances of the Rough Sets Theory*. Norwell, MA: Kluwer, 1992, pp. 331–362.

[11] C. Hor and P. Crossley, "Analysis of substation data for knowledge extraction," presented at the IEEE Power Eng. Soc. General Meeting, San Francisco, CA, Jun. 2–16, 2005.

[12] J. Komorowski, Z. Pawlak, L. Polkowskis, and A. Skowron, "Rough sets: A tutorial," in *Rough Fuzzy Hybridization—A New Trend in Decision Making*, Singapore, 1999.

[13] A. Øhrn, "Discernibility and Rough Sets in Medicine: Tools and Applications," Ph.D. dissertation, Dept. Comput. Sci. Inform. Sci., Norwegian Univ. Sci. Technol., Trondheim, Norway, 2000.

[14] T. Hvidsten, "Fault diagnosis in rotating machinery using rough set theory and ROSETTA," Dept. Comput. Sci. Inform. Sci., Norwegian Univ. Sci. Technol., Trondheim, Norway, 1999, Tech. Rep.

[15] M. Bjanger, "Vibration analysis in rotating machinery using rough set theory and ROSETTA," Dept. Comput. Sci. Inform. Sci., Norwegian Univ. Sci. Technol., Trondheim, Norway, 1999, Tech. Rep.

[16] Ø. Aasheim and H. Solheim, "Rough sets as a framework for data mining," Knowl. Syst. Group, Faculty of Comput. Syst. Telematics, Norwegian Univ. Sci. Technol., Trondheim, Norway, 1996, Tech. Rep.

[17] J. Quinlan, "Generating production rules from decision trees," in *Proc. IJCAI*, 1987, pp. 304–307.

Ching-Lai Hor (M'01) received the B.Eng. (Hons.) degree in electrical and electronic engineering from the University of Manchester Institute of Science and Technology (UMIST), Manchester, U.K., in 1997 and the Ph.D. degree from Queen's University of Belfast, Belfast, U.K., in 2004.

He was an Electrical Engineer with ALSTOM Power from 1997 to 2000 and later as a Postdoctoral Researcher with the Center for Renewable Energy Systems Technology (CREST), Loughborough University. In late 2006, he was appointed Lecturer in Renewable Energy at the University of Exeter in Cornwall, Cornwall, U.K.

Dr. Hor is a member of the Institution of Electrical Engineers (IEE).

Peter A. Crossley (M'95) received the B.Sc. degree from the University of Manchester Institute of Science and Technology (UMIST), Manchester, U.K., in 1977 and the Ph.D. degree from the University of Cambridge, Cambridge, U.K., in 1983.

Currently, he is the Professor of Power Systems at the University of Manchester. He has been involved in various research projects on the applications of global positioning systems (GPS) in electrical networks and the technical problems associated with the connection and transport of electrical energy. He joined Manchester in 2006 after four years at Queen's University of Belfast, 11 years with UMIST, and nine years with GEC/ALSTHOM.

Dr. Crossley is an active member of various CIGRE; IEEE; and Institution of Electrical Engineers (IEE), U.K., committees on protection.

Simon J. Watson (M'05) received the B.Sc. degree in physics from Imperial College, London, U.K., in 1987 and the Ph.D. degree from Edinburgh University, Edinburgh, U.K., in 1990.

He was in the field of renewable energy research in conjunction with power systems at the Rutherford Appleton Laboratory, Oxfordshire, U.K., until 1999. He was then with Good Energy, the green electricity supply company that provides electricity sources from renewable energy generation to domestic and small commercial customers. In 2001, he was appointed Senior Lecturer with the Center for Renewable Energy Systems Technology (CREST), Loughborough University, Loughborough, U.K.