# Reasoning Applied to Substation Data Extracted from Relays and Circuit Breakers

C.L. Hor*
Centre for Renewable Energy Systems Technology
Loughborough University, UK
c.hor@lboro.ac.uk

P.A. Crossley
Electric Power and Energy Systems,
Queen's University of Belfast, UK
p.crossley@ee.qub.ac.uk

**Abstract - The paper describes how rough set theory can extract decision rules from the data acquired by relays and circuit breakers. The methodology correctly estimates the fault section and can help operators in their decision making process. Developing a knowledge based diagnostic system is always costly and time consuming. Extra and superfluous conditions in a knowledge base may lead to serious inconveniences especially in rules maintenance. Reducing the size of the knowledge base and improving the quality of knowledge acquisition, benefits the operating behaviour of the power system. The proposed technique not only induces the decision rules, it also reduces the size of the knowledge base without the loss of useful information. Results can be used by an expert system or a neural network to generate supervisory automation and to support operators during an emergency situation. The former includes the generation of HMI alarms and the latter, the diagnosis of the type and cause of the event and suggestions for network restoration and post emergency repair. The PSCAD/EMTDC simulator has been used to investigate the effect of faults and switching actions 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.**

***Keywords - fault section estimation, relays, circuit breakers, rough sets, discernibility, rules discovery, voting algorithm***

## 1 Introduction

THE amount of data captured within a substation has increased significantly over recent years and human inspection and interpretation may no longer be feasible [1] [2]. Each data record contains a large number of parameters but only some carry the information that is required [3]. In many cases, operators find themselves having only a vague idea of which parameters are important for their analysis. The essential consideration of achieving highly recallable information is to determine the significant attributes in the dataset by filtering the unimportant attributes without losing useful information [4]. Unlike the papers [5] that make use of the supervised rough classification to reduce the overwhelming messages arriving at the control centre for emergency support, this paper presents using the same approach to rule induction for fault diagnosis and online decision support which is simple, robust and consistent. Extracted rules can help domain experts gain insight into the relationships between decision variables so that they can build a more effective knowledge base system. The rules will also be validated on their classification performance using the voting algorithm before they are verified by domain experts.

## 2 Decision System

A decision system contains a set of pre-classified events received from relays. It can be perceived as a two-dimensional data table with a set of data (element of U) represented by rows. Each row corresponds to an event and each column represents an attribute. The decision system can be formulated as $\mathcal{D} = \{U, C \cup D\}$ where universe ***U*** represents a set of time events. ***C*** defines a set of condition attributes i.e. observations and ***D*** is a decision attribute that contains pre-classified events. Any combination of the values for the decision attribute in D is represented by a distinct value for *d*. Thus, D = $\{d\}$.

To overcome the lack of real data for analysis, a typical 132/11kV substation model, as given in Figure 1, was developed [6]. The directional relays at R5 and R6 also include non-directional time graded earth fault elements. This is necessary to protect the 11kV busbar and provide backup for the 11kV feeders [7]. Relays R1, R2, R3 and R4 all gave an identical pattern for the fault F1 and therefore these relays are regarded as one and labelled as "Rx" in which x = $\{1, 2, 3, 4\}$ (See Table 1). Vx and Ix represent the three phase voltages and currents respectively. Similarly, breakers BRK6 and BRK8 are regarded as one and labelled "BZ3". Due to the lack of space, the time events in Table 1 are not displayed. The normal operating voltage range (N) is typically from 0.90 to 1.10 p.u. of the nominal value. Lower than 0.90p.u., the voltage is considered as Low (L) and above 1.10p.u., it is high (H). As the current varies significantly more than the voltage, a wider range of threshold is used. The nominal current range (N) is considered to be between 0.50 and 1.50p.u, meaning that if the current is lower than 0.50pu, it is low (L) and if it is higher than 1.50p.u., it is high (H). The current H1 indicates that it is flowing in the same direction that would trigger the directional relays. $d_1$ indicates the state classifications. Normal (**N**) indicates that all the constraints and loads are satisfied, i.e. the voltages and currents are nominal. Alert (**A**) indicates at least one current is high and the voltages are nominal, or the currents are nominal but at least one voltage is abnormal. Emergency (**E**) indicates at least two physical operating limits are violated (e.g. under voltages and over currents). Safe (**S**) is when those parts of the power system that remain are operating normally, but one or more loads are not satisfied after a breaker has opened [8]. $d_2$ is used to capture the breaker information. $d_2 = 1$ indicates that a breaker has opened and the respective line has been disconnected. R9, R10, R11, R12 are excluded from Table 1 since these unit protection relays do not contribute to this fault (F1) analysis.

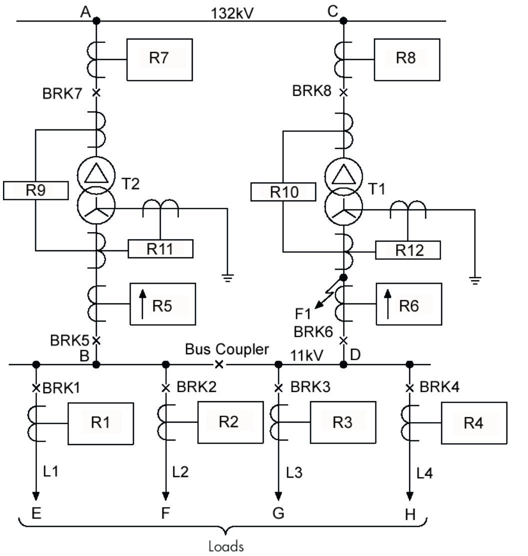


**Figure 1:** 132/11kV Substation model

Table 1 displays the voltage and current patterns captured by the relays in the event of fault F1. The initial data set is too large to include in the paper, hence only change of state is presented.

| **Rx** | | **R5** | | **R6** | | **R7** | | **R8** | | **BRK** | ***d*** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $V_x$ | $I_x$ | $V_5$ | $I_5$ | $V_6$ | $I_6$ | $V_7$ | $I_7$ | $V_8$ | $I_8$ | BZ3 | $d_1 d_2$ |
| N | N | N | N | N | N | N | N | N | N | 0 | N 0 |
| L | N | N | N | N | N | N | N | N | N | 0 | A 0 |
| L | N | L | N | L | N | N | N | N | N | 0 | A 0 |
| L | N | L | H | L | N | N | H | N | H | 0 | E 0 |
| L | N | L | H | L | H1 | N | H | N | H | 0 | E 0 |
| L | L | L | H | L | H1 | N | H | N | H | 0 | E 0 |
| L | L | L | H | L | H1 | N | H | N | H | 1 | E 1 |
| L | N | L | H | L | H1 | N | H | N | H | 1 | E 1 |
| L | N | L | H | L | L | N | H | L | N | 1 | E 1 |
| L | N | L | N | L | L | N | N | L | L | 1 | A 1 |
| N | N | L | N | L | L | N | N | L | L | 1 | A 1 |
| N | N | N | N | L | L | N | N | L | L | 1 | S 1 |

**Table 1:** Decision system

# 3 Rough Sets

### *3.1 Approximations*

The concept of power system states cannot always be defined in a crisp manner using the data collected in a substation. This is where the notion of rough set emerges. Rough set hinges on two basic concepts i.e. the lower and upper approximation. The former indicates the elements that doubtlessly belong to the set whereas the later indicates the elements that possibly belong to the set. The fundamentals of rough set theory for *decision system* are excluded in this paper for they have already been made available in the paper [5]. Nevertheless, two examples are given to demonstrate how rough sets and the discernibility matrix are used to compute the **reducts**[1]. A relative discernibility matrix is applied to the minimal attribute set to look for the **core**[2] before any rules are extracted.

### *3.2 Discernibility Matrix*

In a decision system, the same cases could occur several times or some attributes could be superfluous. Those unnecessary attributes should be eliminated using a discernibility matrix. It is a symmetric $n \times n$ matrix where *n* denotes the number of elementary sets [9]. Assume that the attribute B $\subseteq$ A and the decision table is represented as $\mathcal{D} = (\mathrm{U}, \mathrm{B} \cup \{d\})$. The discernibility matrix, $M^d(B)$ can thus be formulated as follows: -

$$M^d(B) = \left\{m_B^d(x_i, x_j)\right\}_{n \times n},$$
$$m_B^d(x_i, x_j) \begin{cases} \emptyset & if\ \forall\, d \in D\, [d(x_i) = d(x_j)] \\ \{r \in B : r(x_i) \neq r(x_j)\} & \\ & if\ \exists d \in D\, [d(x_i) \neq d(x_j)] \end{cases} \tag{1}$$

where $i, j = \{1, ..., n\}$ and $n = |\mathrm{U/IND\,(B)}|$

The notion $r(x)$ denotes the set of decisions for a given class $x \in \mathrm{U/IND\,(B)}$. The entry $m_B^d(x_i, x_j)$ in the discernibility matrix is the set of all (condition) attributes from B that classify events $x_i$ and $x_j$ into different classes in U/IND(B) if $r(x_i) \neq r(x_j)$. Empty set $\emptyset$ denotes that this case does not need to be considered. All the disjuncts of the minimal disjunctive form of this function define the reducts of B [10].

### *3.3 Discernibility Functions*

After the discernibility matrix has been created, the discernibility function can be defined. 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. Given a decision system $\mathcal{D} = (\mathrm{U}, \mathrm{B} \cup \{d\})$, the discernibility function is: -

$$f_B^d(x_i) = \bigwedge \left\{ \bigvee \bar{m}_B^d(x_i, x_j) \ : \ 1 \leq j \leq i \leq n \right\} \tag{2}$$

where $(\bigvee)$ and $(\bigwedge)$ are the disjunction and conjunction operators. $n = |\mathrm{U/IND\,(B)}|$ and $\bigvee \bar{m}_B^d(x_i, x_j)$ 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 $\emptyset$ [10].

The decision relative discernibility function of B can be constructed to discern an event belonging to another class such as for an event class $x_k = (1 \leq k \leq n)$ over attributes B, it can be represented in Equation 3.

$$f(x_k, B) = \bigwedge \left\{ \bigvee \bar{m}_B^d(x_k, x_j) \ : \ 1 \leq j \leq n \right\} \tag{3}$$

This function computes the minimal set of attributes B necessary to distinguish $x_k$ from other event classes defined by B [10].

## 4 Rule Accuracy and Assessment

A decision rule can be denoted $\alpha \rightarrow \beta$, read as "if $\alpha$ then $\beta$". The pattern $\alpha$ is called the rule's antecedence while the pattern $\beta$ is called the rule's consequence. Three units of measure shown below can be used to evaluate the quality of a given decision rule [11]:

1. *Support*: the number of events that possesses both property $\alpha$ then $\beta$.

[1] REDUCT is a reduced set of relations that ensures the same quality approximation as the whole set of attributes.
[2] CORE is the set of relations occurring in every reduct, i.e. the set of all indispensable relations to characterize the equivalence relation.

2. *Accuracy*: A decision rule $\alpha \rightarrow \beta$ may only reveal partially the overall picture of the derived decision system. Given pattern $\alpha$, the probability of the conclusion $\beta$ can be assessed by measuring how trustworthy the rule is in drawing the conclusion $\beta$ on the basis of evidence $\alpha$ .

$$Accuracy\,(\alpha \rightarrow \beta) = \frac{support\,(\alpha \cdot \beta)}{support\,(\alpha)} \quad (4)$$

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 $(\alpha \rightarrow \beta)$ is required in order to measure how well the evidence $\alpha$ describes the decision class. It can be defined through $\beta$: -

$$Coverage\,(\alpha \rightarrow \beta) = \frac{support\,(\alpha \cdot \beta)}{support\,(\beta)} \quad (5)$$

## 5 Voting Algorithm

There are various ways of classifying events using rule sets, and the voting algorithm can be used to resolve the conflicts and rank the predicted outcomes. This works reasonably well for rule-based classification. Let RUL denotes an unordered set of decision rules. The voting process among the rules that fire, 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 [11]:

1. The set of rules RUL searches for applicable rules RUL($x$) that match the attributes of event $x$ (i.e. rules that *fire*) in which RUL($x$) $\subseteq$ RUL.

2. If no rule is found i.e. 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: -

   - *Casting the votes*: Let a rule $r \in$ RUL($x$) casts as many votes, *votes(r)* in favour of its outcomes associated with the support counts as given by Equation (6):-

   $$votes(r) = |\;||\alpha \cap \beta||\;| \quad (6)$$

   - *Compute a normalisation factor, norm(x)*. The normalisation factor is computed as the total number of votes cast and only serves as a scaling factor in Equation (7): -

   $$norm(x) = \sum_{r \in RUL(x)} votes\,(r_i) \quad (7)$$

   - *Certainty Coefficient*: The votes from all the decision rules $\beta$ are accumulated before they are divided by the normalisation factor norm($x$) to yield a numerical certainty coefficient. Certainty$(x, \beta)$ for each decision class is given in Equation (8): -

   $$Certainty\,(x, \beta) = \left(\frac{votes\,(\beta)}{norm\,(x)}\right) \quad (8)$$

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

## 6 Example I

This example considers a scenario which involves a Fault (F1) on the 11kV transformer T1 feeder given in Figure 1. The fault results in the operation of the directional relay R6, the tripping of circuit breakers BRK6 and BRK8 and the isolation of the transformer T1. The decision system in Table 1 is transformed into a discernibility matrix shown in Table 2 using Equation 1.

| | 1 | 2 | 3 | $\cdots$ | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|
| 1 | $\emptyset$ | | | | | | |
| 2 | x | $\emptyset$ | | | | | |
| 3 | x,5,6 | $\emptyset$ | $\emptyset$ | | | | |
| 4 | x-8 | 5-8 | 5,7,8 | $\cdots$ | | | |
| 5 | x-8 | 5-8 | 5-8 | $\cdots$ | | | |
| 6 | x-8 | x-8 | x-8 | $\cdots$ | | | |
| 7 | x-8B | x-8B | x-8B | $\cdots$ | | | |
| 8 | x-8B | 5-8B | 5-8B | $\cdots$ | | | |
| 9 | x-8B | 5-8B | 5-8B | $\cdots$ | | | |
| 10 | x,5,6,8,B | 5,6,8,B | 6,8,B | $\cdots$ | $\emptyset$ | | |
| 11 | 5,6,8,B | x,5,6,8,B | x,6,8,B | $\cdots$ | $\emptyset$ | $\emptyset$ | |
| 12 | 6,8,B | x,6,8,B | x,5,6,8,B | $\cdots$ | x,5 | 5 | $\emptyset$ |

**Table 2:** Discernibility matrix

$$\text{where :}\quad \begin{array}{lcl} \{\mathrm{x}-8\} & = & \{\mathrm{x}, 5, 6, 7, 8\} \\ \{\mathrm{x}-8\mathrm{B}\} & = & \{\mathrm{x}, 5, 6, 7, 8, \mathrm{BZ3}\} \\ \{5-8\} & = & \{5, 6, 7, 8\} \\ \{5-8\mathrm{B}\} & = & \{5, 6, 7, 8, \mathrm{BZ3}\} \end{array}$$

Based on the discernibility functions derived from each column in Table 2 using Equation 2, the final discernibility function computed is thus: $f(\mathcal{D}) = \mathrm{Rx} \cdot \mathrm{R5} \cdot \mathrm{BZ3}$, in which the form '$\cdot$' refers as the operator of conjunction ($\bigwedge$). As Rx = $\{\mathrm{R1}, \mathrm{R2}, \mathrm{R3}, \mathrm{R4}\}$ and BZ3 = $\{\mathrm{BRK6}, \mathrm{BRK8}\}$, a total of 8 reducts can be generated. Depending on the data availability, either one of these reducts can be used to classify the events. In other word, if there are some missing sources 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 3.

| Rule | Rx | | R5 | | BRK | $d$ | Support |
|---|---|---|---|---|---|---|---|
| No. | $\mathrm{V_x}$ | $\mathrm{I_x}$ | $\mathrm{V_5}$ | $\mathrm{I_5}$ | BZ3 | $d_1 d_2$ | count |
| 1 | N | N | N | N | 0 | N 0 | 1 |
| 2 | L | N | N | N | 0 | A 0 | 1 |
| 3 | L | N | L | N | 0 | A 0 | 1 |
| 4 | L | N | L | H | 0 | E 0 | 2 |
| 5 | L | L | L | H | 0 | E 0 | 1 |
| 6 | L | L | L | H | 1 | E 1 | 1 |
| 7 | L | N | L | H | 1 | E 1 | 2 |
| 8 | L | N | L | N | 1 | A 1 | 1 |
| 9 | N | N | L | N | 1 | A 1 | 1 |
| 10 | N | N | N | N | 1 | S 1 | 1 |

**Table 3:** Reduct table

### 6.1 *Quality of rule measure*

The quality of rules from Table 3 can be assessed based on the unit of measure i.e. RHS and LHS support, accuracy coverage and length in Table 4. The **LHS** ("*left hand side*") support signifies how many events are in the data set. The **RHS** ("*right hand side*") support signifies how many events in the data set that match the if-part and have the decision value of the then-part. For an inconsistent rule, then-part shall consist of several decisions. *Accuracy* and *coverage* are computed from the support counts using Equation 4 and 5. Since there is no inconsistency in the decision system, the accuracy of rules are thus 1.0. *Length* indicates the number of attributes in the LHS or RHS; LHS = 3 (Rx, R5, BZ3) and RHS = 1.

| 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 |

**Table 4:** Quality of rule measure

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

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

**Table 5:** Relative discernibility matrix

### 6.2 *Relative discernibility functions*

Table 3 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 given in Equation 3. It is based on the relative discernibility matrix constructed for the subspace $\{\mathrm{Rx, R5, BZ3}\}$ as shown in Table 5. Because of the space constraint in Table 5, let Rx $= \{\mathrm{V_x, I_x}\}$ and R5 $= \{\mathrm{V_5, I_5}\}$. Voltage and current attributes in each relay are considered separately rather than treating them as one unit as in Table 2. In each column of Table 5, the relative discernibility functions are computed. For example, to construct $f(1,\ \mathrm{B})$ where $\mathrm{B} \subseteq \mathrm{A}$, all sets of attributes from column 1 are summed using the absorption law, similarly for $f(2,\ \mathrm{B})$ with all sets of attributes from the column 2 and so on.

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

The form '·' refers as the operator of conjunction ($\bigwedge$) and '+' as the operator of disjunction ($\bigvee$). The result for $f$(2,B) indicates that there are two rules to classify the abnormal state. The first rule requires the attributes $\{\mathrm{V_x, V_5, I_5}\}$ whereas the second rule requires the attributes $\{\mathrm{V_x, V_5, BZ3}\}$. The relative discernibility functions are converted into 16 decision rules listed in Table 6. Among the rules, one of the two are actually redundant i.e. 4 and $5^{(1)}$, $6^{(1)}$ and 7. They are filtered out leaving only 14 applicable rules. The set of rules in Table 6 is categorised into 5 different classes according to their outcomes:

1. **ABNORMAL A0**
   **Rule 1:** $\mathrm{V_x}$(L), $\mathrm{V_5}$(N), $\mathrm{I_5}$(N) Z1
   **Rule 2:** $\mathrm{V_x}$(L), $\mathrm{V_5}$(N), BZ3(0) Z1
   **Rule 3:** $\mathrm{V_x}$(L), $\mathrm{I_5}$(N), BZ3(0) Z1
   **Rule 4:** $\mathrm{V_5}$(L), $\mathrm{I_5}$(N), BZ3(0) Z25
   The system behaves abnormally and is at high alert. Zone Z1 and Z2 both experience voltage sags.

   Referring to Figure 1, the substation can be divided into four main protection zones. Zone 1 represents the protection zones of R1, R2, R3 and R4. Zone 2 the zones of R5, R7, R9 and R11. Zone 3 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 5.

2. **ABNORMAL A1**
   **Rule 5:** $\mathrm{V_x}$(L), $\mathrm{I_5}$(N), BZ3(1) Z1 & Z3
   **Rule 6:** $\mathrm{V_5}$(L), $\mathrm{I_5}$(N), BZ3(1) Z25 & Z3
   ***Rule 7:*** $\mathrm{V_x}$(N), $\mathrm{V_5}$(L), $\mathrm{I_5}$(N) Z25
   **Rule 8:** $\mathrm{V_x}$(N), $\mathrm{V_5}$(L), BZ3(1) Z25 & Z3
   The system is recovering. Protection at Zone 3 has responded. The situation is under control but not safe.

3. **EMERGENCY E0**
   **Rule 9:** $\mathrm{I_5}$(H), BZ3(0), Z25
   ***Rule 10:*** $\mathrm{I_x}$(L), BZ3(0), Z1
   The system is unstable and an urgent action is required. Protection has not yet responded.

4. **EMERGENCY E1**
   ***Rule 11:*** $\mathrm{I_x}$(L), BZ3(1), Z1 & Z3
   ***Rule 12:*** $\mathrm{I_5}$(L), BZ3(1), Z25 & Z3
   The system is still unstable. Protection at Zone 3 has responded. The fault is isolated to Zone 3.

5. **SAFE S1**
   **Rule 13:** $\mathrm{V_5}$(N), BZ3(1), 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, 12 may have to be modified as it does not clearly justify the status. This does not mean that the rules extraction is inaccurate, simply because the data set does not contain adequate information to classify the events.

| Rules | Rx | | R5 | | BRK | $d$ | Sup.1 | Sup.2 |
|---|---|---|---|---|---|---|---|---|
| No. | $V_x$ | $I_x$ | $V_5$ | $I_5$ | BZ3 | $d_1 d_2$ | 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 |

**Table 6:** Core table

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

Different set of decisions can be fired based on the rule's consequence(s). A lookup table can be used to retrieve the mapping between the input values and the rule's consequence(s) for each scenario. If the fault symptom matches the list of the rules (facts) given above, a fault in Zone Z36 is concluded. The example shows that the approach is capable of inducing the decision rules from a substation database, even though the data set may contain only a reasonable quality of information.

### 6.3 *Voting results*

The rules derived from the reducts should be assessed on its classification performance, readability and usefulness before they can be used effectively for online diagnosis. Table 7 illustrates the results computed by the voting algorithm. Assume that only the rules presented with $V_5 = L$ and $I_5 = N$ are fired, the voting algorithm based on the Support Count Index 1 concluded the ABNORMAL decision (combining the result of A0 = $4/9$ and A1 = $5/9$). The support count for the case $V_5 = L$ or $I_5 = N$ that equal to the outcome A0 is 4, whereas the total support count for the case $V_5 = L$ or $I_5 = N$ regardless any outcome is 9. The same procedure applies to the A1 in which the support count for the case $V_5 = L$ or $I_5 = N$ that equal to the outcome A1 is equal to 5. With the set of given rules, the most likely decision value is thus an ABNORMAL state. Considering which abnormal states will be fired, it shall be A1. Now, assume that only rules presented with $V_x = L$ and $V_5 = L$ and $I_5 = H$ are fired. 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 which is 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 by some experts. The reason for this conflict is caused by the inadequate information in the small data set in Table 1. As the result, the rule coverage is limited particularly on the emergency period. To support our explanation, we apply the Support Count Index 2 based on a more complete data set that contains a three phase currents and a 3-phase voltage. The same procedure is repeated and this time, the emergency state is chosen which can be seen in Table 7 with the certainty coefficients computed for each decision class. The suggestion from the voting result should be left to operators/experts to decide the necessary actions.

| **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 |

**Table 7:** Accumulating the casted votes for all rules that fires

### 6.4 *Classifier performance*

For assessing the classifier performance, the data set is divided into a training set and a test set. The training set is a set of examples used for learning that is to fit the parameters, whereas the test set 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 test set. If the rules cannot classify the events in the test set satisfactorily, the rules must be notified to the user and refined to suit the real application.

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 which was tested on every split of training set for a total for 4 runs. The splits are used to avoid results based on rules that were generated for a particular selection of events. This makes the results more reliable and independent on one particular selection of events.

Table 8 and Table 9 show that we have achieved a **100%** in the accuracy of classification for the 10% and 30% test set. Finally, when 50% of the data is used, the accuracy dropped to **94.6%**. The overall results have proven that the extracted rules have a very high and successful classification rate.

| 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 8:** Classifier result using the 90% training set and 10% test set

| 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 9:** Classifier result using the 70% training set and 30% test set

| 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** |

**Table 10:** Classifier result using the 50% training set and 50% test set

## 7 Example II

Table 11 and Table 12 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 (See Figure 1). Bx = the breaker x in which x = $\{1, 2, 3, 4\}$. Similar to BZ3, BRK5 and BRK7 are regarded as one and labelled "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. The reason of acquiring the auxiliary contacts is to capture the right information in case the protection system has failed/maloperated. The current H1 indicates that it is flowing in the same direction that would trigger the directional relays. Ix $= \{I_1, I_2, I_3, I_4\}$ since all the load currents have a similar patterns. Going through the same procedure as described in Section 2, the results obtained based on Table 11 are as follows:

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

**Table 13:** Rules generated for various fault scenarios in the substation

Combining the information from Table 12 and 13, six concise decision rules can be obtained which can be interpreted as follows: -

**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 protections.

**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 given example is small and incomplete. Therefore, some of these extracted rules may look a little bit oversimplified. This likely to happen when the data set 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 look also predictable for a small substation like in Figure 1. However, considering a larger substation or a complex power network with a large number of protection system(s), extracting rules from such circumstance may consume a lot of time and manpower. As such, this method will be useful to power utilities for exploiting substation rules. It also help reducing the size of conventional rule base system by eliminating the extra and superfluous conditions 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 concerning about 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.

## 8 Conclusion

This paper suggests the use of a novel, structured method to reason and extract implicit knowledge from operational data derived from relays and circuit breakers. The proposed analytical method has been used to identify underlying data relationship and simplified logic based rules that can be used to identify or classify the fault section and abnormal events.

The theoretical approach taken is simple but robust and the resulting method has shown promises for eventual application in the power system engineering domain. The methodology is more attractive than some other techniques like Bayesian approach because no assumption about the independence of the attributes are necessary nor is any background knowledge about the data [12]. Therefore, a set of training data of reasonable quality is needed.

| 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 11:** 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 | 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 |

**Table 12:** List of switching actions with estimated protection zones for various fault scenarios

Though decision trees have been used successfully in ID3 and C4.5, compare to rule set generated by rough set theory, it remains questionable whether decision trees can be described as knowledge, no matter how well they function [13]. Their performance can also be affected in the presence of missing values in the test data set which is less likely the case for rough set theory.

Our rules extraction and subsequent classification can be performed without the presence of an expert. However, experts may still have to perform the final check before these rules are used in the real time application. The technique simplifies the rule generation (knowledge acquisition) and reduces the time and manpower 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 induced from the data sets and from these simplified examples, the results shown look promising.

## REFERENCES


[1] M. Eby. Don't let data overload stop you. *Transmission & Distribution World*, May 1 1999.

[2] W. Ackerman. The Impact of IEDs on the Design of Systems used for Operation and Control of Power Systems. In *the Proceeding of the IEE $5^{th}$ International Conference on Power System Management and Control (PSMC)*, London, UK, April 2002.

[3] L. Smith. Analysis of Substation Data. Technical report, Report by Working Group I-19 to the Relaying Practices Subcommittee of the Power System Relaying Committee, IEEE Power Engineering Society, 2002.

[4] X. Hu. *Knowledge Discovery In Databases: An Attribute-Oriented Rough Set Approach.* PhD thesis, Department of Computer Science, University of Regina, June 1995.

[5] C. Hor, P. Crossley. Extracting Knowledge from Substations for Decision Support. *IEEE Transactions on Power Delivery*, 20(2), pp.595-602, April 2005.

[6] C. Hor, A. Shafiu, P. Crossley, and F. Dunand. Modelling a substation in a distribution network: real time data generation for knowledge extraction. *IEEE Power Engineering Society Summer Meeting*, Chicago, Illinois, USA, July 21-25, 2002.

[7] C. Hor, K. Kangvansaichol, P. Crossley, and A. Shafiu. Relay modelling for protection studies. *IEEE Bologna Power Tech 2003 Conference*, Bologna, Italy, June 23-26, 2003.

[8] G. Gross, A. Bose, C. DeMarco, M. Pai, J. Thorp, and P. Varaiya. Real Time Security Monitoring and Control of Power Systems. Technical report, The Consortium for Electric Reliability Technology Solutions (CERTS) Grid of the Future White Paper, December 1999.

[9] A. Skowron and C. Rauszer. *The discernibility matrices and functions in information systems.* Intelligent Decision Support - Handbook of Applications and Advances of the Rough Sets Theory. pp331-362, Kluwer Academic Publishers, Dordrecht, New Haven, 1992.

[10] J. Komorowski, Z. Pawlak, L. Polkowskis, and A. Skowron. *Rough Sets: A Tutorial.* In: Rough Fuzzy Hybridization – A New Trend in Decision Making, Springer Verlag Publisher, Singapore, 1999.

[11] A. Øhrn. *Discernibility and Rough Sets in Medicine: Tools and Applications.* PhD thesis, Department of Computer Science and Information Science, Norwegian University of Science and Technology, Trondheim, Norway, February 2000.

[12] Ø. Aasheim and H. Solheim. Rough Sets as a Framework for Data Mining. Technical report, Knowledge Systems Group, Faculty of Computer Systems and Telematics, The Norwegian University of Science and Technology, Trondheim, Norway, May 1996.

[13] J. Quinlan. Generating Production Rules from Decision Trees. Proc. IJCAI-87,304-307 (1987).