Journal of Applied Sciences Research, 5(10): 1698-1708, 2009

Making Implicit Knowledge of Distance Protective Relay Operations and Fault Characteristics Explicit via Rough Set Based Discernibility Relationship

[1]Mohammad Lutfi Othman, [1]Ishak Aris, [1]Senan Mahmod Abdullah,
[2]Mohammad Liakot Ali Mohammad Abu Bakar, [3]Mohammad Ridzal Othman

[1]Department of Electrical and Electronic Engineering,
Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.
[2]Institute of Information and Communication Technology, Bangladesh University of Engineering and Technology, Dhaka-1000, Bangladesh.
[3]Engineering Department, Tenaga Nasional Berhad, Crystal Plaza, Jalan 223/51 A,
46100 Petaling Jaya, Selangor, Malaysia.

Abstract: This paper discusses the novel application of the discernibility concept inherent in rough set theory in making explicit of the implicit knowledge of distance protective relay operations and fault characteristics that are hidden away in the recorded relay event report. A rough-set-based data mining strategy is formulated to analyze the relay trip assertion, impedance element activation, and fault characteristics of distance relay decision system. Using rough set theory, the uncertainty and vagueness in the relay event report can be resolved using the concepts of discernibility, elementary sets and set approximations. Nowadays protection engineers are suffering from very complex implementations of protection system analysis due to massive quantities of data coming from diverse points of intelligent electronic devices (IEDs such as digital protective relays, digital fault recorders, SCADA's remote terminal units, sequence of event recorders, circuit breakers, fault locators and IEDs specially used for variety of monitoring and control applications). To help the protection engineers come to term with the crucial necessity and benefit of protection system analysis without the arduous dealing of overwhelming data, using recorded data resident in digital protective relays alone in an automated approach called knowledge discovery in database (KDD) is certainly of an immense help in their protection operation analysis tasks. Digital protective relay, instead of a host of other intelligent electronic devices, is the only device for analysis in this work because it sufficiently provides virtually most attributes needed for data mining process in KDD. Unlike some artificial intelligence aproaches like artificial nueral network and decision tree in which the data mining analysis is "population-based" and single since it is common to the entire population of training data set, the rough set approach adopts an "individually-event-based" paradigm in which detailed time tracking analysis of relay operation has been successfully performed.

Key words: Decision system, distance protection, digital protective relay, data mining, knowledge discovery in database, rough set theory.

INTRODUCTION

Modern microprocessor-based intelligent electronic devices (IEDs) provide a large amount of data about power system operations and protection schemes. Using suitable data mining and analysis techniques, relevant information from the stored event and fault reports and oscillography and setting files can be analyzed for the purpose of improving utility protection quality.

In doing so, protection engineers are nowadays suffering from data overload: more data than can be processed and assimilated for knowledge discovery and decision support in the time available[1]. The massive quantities of data coming from diverse points of IEDs (such as digital protective relays (DPRs), digital fault recorders (DFRs), SCADA's remote terminal units (RTUs), sequence of event recorders (SERs), circuit breakers (CBs), fault locators (FLs) and IEDs specially used for variety of monitoring and control applications) in vast array of implementations make analysis of

Corresponding Author: Mohammad Lutfi Othman, Department of Electrical and Electronic Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia.
Tel: 603 89466325, Fax: 603 89466327, H/P: 60192755209
E-mail: lutfi@eng.upm.edu.my

protection system analysis a very complex area of study. Automatically collecting, filing, and managing large volumes of IED data is a monumental task[2]. To help the protection engineers come to term with the crucial necessity and benefit of protection operation analysis without the uneasiness of dealing with overwhelming data, using recorded data resident in DPRs alone is certainly of an immense help in their analysis tasks.

Digital Protective Relay Event Report as the Source of Protection Analysis: Due to the fact that problem involving protection system is usually nonlinear, wide-scale and an amalgamation of integral subsystems, various artificial intelligence techniques have been extensively explored and successfully implemented with[3]. In the last twenty years or so, the automated protection analysis research has advanced with new developments being pursued charting into two primary courses, which are:

- application of intelligent system technologies with artificial intelligence (AI) methods as the main thrusts and
- utilization of intelligent electronic devices' (IEDs) recorded data, inseparably[1].

Significant research has been carried out in the area of fault and protection system diagnosis in power system using various AI such artificial nueral network [3], event tree analysis[4], expert system[5], model-based reasoning[6,7] and fuzzy logic[8,1,9,10].

Nevertheless, the primary aim of integrating intelligent techniques with various IEDs is mainly for "fault response analysis" (sometimes called "postmortem fault analysis" and "protection performance analysis") rather than detailed validation and diagnosis of digital protective relay behavior analysis using data from the relays alone. The protection operation analysis approaches have been geared towards protection system of a specific scale of power system such as that of a distribution system or a specific span of transmission system involving collectively a set of protection devices or components [11].

Modern DPRs provide a wide spread value of information, internal and external to it. It has become the most preferred recording system in this era of microprocessor advancement and innovation due to its reasonable price, reliability, multifunctionality and flexibility[12]. For reasonably detailed analysis of protection operation in this work, protective relay is regarded as the main source of data especially the operational ones such as time tagged currents and voltages (than the non-operational) (Hor, 2006). Only protective relays can provide the relay target data which is obviously absent or not fully available in SERs, DFRs and SCADA RTUs[13].

Thus, the study in this paper will focus on the protective relay operation analysis using an intelligent technique of computation called rough set theory. The crux of rough set based data analysis is an information system, a tabulated data set of knowledge representation system, whose columns are labeled by attributes, rows are labeled by objects of interest and entries of the table are attribute values[14].

Recently, rough set theory approaches have been successfully applied to power system event analysis using information from any of the IEDs found in the substations to derive useful and comprehensive knowledge of the power system conditions in terms of faults and protection system operations[15,16,17]. Most of these works focus on 'system' rather than 'device' in the protection operation analysis. In view of that, the emphasis of the work in this paper is mainly formulating the rough-set-based data mining strategy of relay event report derived from digital protective relay under analysis and see how this strategy makes detailed device-level relay operation analysis possible.

Rough Set Theory as a Suitable Data Mining Technique on Relay Decision System: Introduced by Pawlak[18] in the early 1980s rough set theory is an alternative tool in intelligent data analysis and data mining that can be employed to handle uncertainty and vagueness[14].

It is usually very difficult to select a group of effective attributes to fully reflect relay behavior because of the highly non-linear nature of relay operation analysis. The selected attributes seldom provide adequate knowledge to accurately map the interclass boundary, making the inter-class boundary usually "rough". In other words, some cases close to the boundary are practically unclassifiable based on the selected attributes. Therefore, the protective relay operation analysis is actually a rough classification problem where there are small overlaps between the different classes. This leads to uncertainty and vagueness in different classes.

Of particular interest in this study of relay analysis is the case where upon fault inception, a protective relay picks it up and provides a common combination of tripping conditions in time sequence succession but having two distinct representations (classifications) in tripping decision: one when trip signal has not been asserted immediately after relay pick-up and the other is when trip signal is asserted, after a preset time delay as set by the protection engineer. Thus, rough set theory suits the bill in resolving this conflict.

Using rough set theory, the above conflict resident in the relay event report is resolved using the concepts of discernibility, elementary sets and set approximation. A rough set is a formal approximation of a crisp set (i.e., conventional set) in terms of a pair of sets which give the *lower* and the *upper* approximation of the original set. The lower and upper approximation sets themselves are crisp sets[19].

On why rough set is chosen instead of other computational intelligence techniques in this study, the following are some of the reasons:

- The use of rough set doesn't need any additional information about data (e.g. basic probability assignment in the Dempster–Shafer theory of evidence; or grades of attribute membership in fuzzy set theory[14,20] and permits, without loss of accuracy, minimizing of knowledge base, represented as a set of decision rules[21].
- Unlike the black box process of artificial nueral network in which the output is blindly trusted although the knowledge is not comprehensible and easily justifiable, the rule discovering process of rough set is intuitively comprehensible that can be interpreted and analyzed for intelligent decision making support[22].
- Unlike artificial nueral network and decision tree in which the model discovered is "population based" and single since it is common to the entire population (training data set), the rough set approach adopts an "individual (data object/event) based" paradigm in which the models (rules) discovered are explicit and multiple[23].

MATERIALS AND METHOD

Researh Methodology in Distance Relay Operation Analysis: Under the concepts of discernibility, elementary sets and set approximation in rough set theory, the following analysis shall be performed:

- Discovering knowledge of relay trip assertion as shown by the flow chart in Fig. 1
- Discovering knowledge of impedance element activation and fault characteristics as shown by the flow chart in Fig. 2.

According to the rough set philosophy, every relay event of the universe of discourse is assumed to be associated with some information (data, knowledge). For instance, in the sequential times of relay events in event report of protective relay operation, the measurands of ac voltages and currents and the state of multifunctional protective elements form information about time (i.e. the behavior or condition of the relay at different time instances). Motivated by rough set concept [14,19], the following are the mathematical basis of rough set theory that can deal with event-based analysis of relay decision system *DT*:

- Relay events characterized by the same information description (descriptor (q,v)) are *indiscernible* (similar) in view of the available information about them.
- Any set of all indiscernible (similar) relay eventss is called *elementary set*.
- An elementary set form basic *granule* (atom) of knowledge about the universe.
- Any union of some elementary sets is referred to as *crisp* (precise) set - otherwise a set is *rough* (imprecise, vague).
  - Rough set has boundary-line cases, i.e., relay events which cannot be classified with certainty as members of the set or of its complement. Boundary-line cases cannot be properly classified by employing the available knowledge.
  - Crisp sets have no boundary-line elements at all.
- Relay events that can be 'seen' only through the information available about them suggests that knowledge has a *granular* structure.
  - Some relay events of interest appear as the same (or indiscernible/similar), i.e., cannot be discerned (differentiated).
- Vague concepts, in contrast to precise concepts, cannot be characterized in terms of information about their elements but are replaced by a pair of precise concepts, i.e.,:
  - *Lower approximation* of the vague concept, consisting of all relay events which surely belong to the concept.
  - *Upper approximation* of the vague concept, containing of all relay events which possibly belong to the concept.
- The difference between the upper and the lower approximation constitute the *boundary region* of the vague concept.

Upper and lower approximations are two basic operations in the rough set theory to be used in protective relay analysis.

Distance Relay Decision System (*DT*): The pre-data-mining decision system *DT* of a PSCAD-modelled distance protective relay, after having been subjected to zone 1 A-G fault, is shown in Table 1. The power system upon which the relay is to protect is a 500kV double-sourced transmission system modeled in the PSCAD/EMTDC[24].

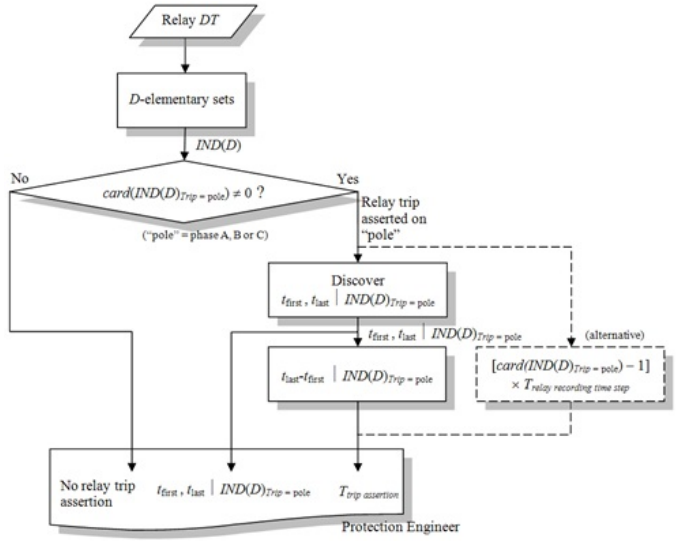

Fig. 1: Flowchart showing the pipelines of discovering the knowledge of relay trip assertion.

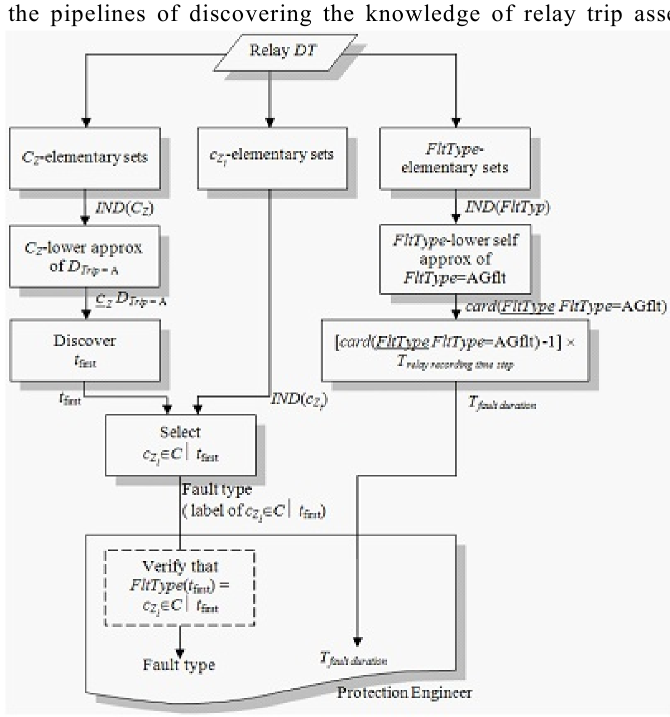

Fig. 2: Flowchart showing the pipelines of discovering knowledge of impedance element activation and fault characteristics

This decision system is a 4-tuple structure
$DT = <U, Q, V, f>$ where,

- U is a finite set of instances/objects (relay events) t_i's, i.e. the *universe* denoted as $U = \{t_1, t_2, t_3, \ldots, t_m\}$.
- $Q = C \cup D$ is a non-empty finite union set of condition and decision *attributes* (condition attributes $c_i \subset C$ denote the internal multifunctional protective elements while decision attribute $d_i \subset D$ denotes the trip output of the relay), such that q: $U \rightarrow V_q$ for every $q \in Q$.
- $V = U_{q \in Q} V_q$, and V_q, is a *domain* (set of values) of the attribute q.
- f: $U \times Q \rightarrow V$ is a total function , called *information function* (alternatively denoted r) such that $f(t,q) \in V_q$ for every $q \in Q$, $t \in U$. Any pair (q,v), where $q \in Q$ and $v \in V_q$, is called *descriptor* in DT

Table. 1: The pre-data-mining DT of distance protective relay subjected to zone 1 A-G fault.

Time code	time	Zab	Zbc	Zca	Zag	Zbg	Zcg	CB52_A	CB52_B	CB52_C	VTmcb	VTSfail_cnfm	CR	pg_PkUp	pp_PkUp	FltType	pp50_Z1	pp50_Z2	pp50_Z3	pp50_Z4	p50_Z1	p50_Z2	p50_Z3	p50_Z4	r50	PSB	Q32	Zload	Q50	Dist_ab	Dist_bc	Dist_ca	Dist_ag	Dist_bg	Dist_cg	PG_Trp	PP_Trp	Trip_CAPerm	Trip_CABlck	WI_Trp	WI_CRTrp	DEFelem_Trp	DEF_WI_CRTrp	Trip_DEF_CA	DEF_WI_Trp	Trip
t_{76}	0.090036	0	0	0	0	0	0	closed	closed	closed	0	0	0	0	0	no fault	0	0	0	0	0	0	0	0	0	0	Fwd	Fwd	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t_{77}	0.091236	0	0	0	0	0	0	closed	closed	closed	0	0	0	0	0	no fault	0	0	0	0	0	0	0	0	0	0	Fwd	Fwd	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t_{78}	0.092437	0	0	0	0	0	0	closed	closed	closed	0	0	0	0	0	no fault	0	0	0	0	0	0	0	0	0	0	Fwd	Fwd	0	0	0	0	0	0											0	0
t_{79}	0.093637	0	0	0	0	0	0	closed	closed	closed	0	0	0	0	0	no fault	0	0	0	0	0	0	0	0	0	0	Fwd	Fwd	0	0	0	0	0	0											0	0
t_{80}	0.094838	0	0	0	0	0	0	closed	closed	closed	0	0	0	0	0	no fault	0	0	0	A	0	A	A	A	0	0	Fwd	Fwd	0	0	0	0	0	0											0	0
t_{81}	0.096038	0	0	0	0	0	0	closed	closed	closed	0	0	0	0	0	no fault	0	A	A	A	A	A	A	A	0	0	Fwd	Fwd	0	0	0	0	0	0											0	0
t_{82}	0.097239	0	0	0	0	0	0	closed	closed	closed	0	0	0	0	0	AGflt	A	A	A	A	A	A	A	A	34	0	Fwd	Fwd	4	0	0	0	0	0											0	0
t_{83}	0.098439	0	0	0	0	0	0	closed	closed	closed	0	0	0	0	0	AGflt	A	A	A	A	A	A	A	A	234	0	Fwd	Fwd	4	0	0	0	0	0											0	0
t_{84}	0.09964	0	0	0	0	0	0	closed	closed	closed	0	0	0	0	0	AGflt	A	A	A	A	A	A	A	A	1234	0	Fwd	Fwd	1234	0	0	0	0	0											0	0
t_{85}	0.10084	0	0	0	0	0	0	closed	closed	closed	0	0	0	0	0	AGflt	A	A	A	A	A	A	A	A	1234	0	Fwd	Fwd	1234	0	0	0	0	0											0	0
t_{86}	0.102041	0	0	0	3	0	0	closed	closed	closed	0	0	0	23	0	AGflt	A	A	A	A	A	A	A	A	1234	0	Fwd	Fwd	1234	0	0	0	23	0											0	0
t_{87}	0.103241	0	0	0	23	0	0	closed	closed	closed	0	0	0	23	0	AGflt	A	A	A	A	A	A	A	A	1234	0	Fwd	Fwd	1234	0	0	0	23	0											0	0
t_{88}	0.104442	0	0	0	23	0	0	closed	closed	closed	0	0	0	23	0	AGflt	A	A	A	A	A	A	A	A	1234	0	Fwd	0	1234	0	0	0	23	0	0	0			0	0	0	0	0	0	0	0
t_{89}	0.105642	0	0	0	23	0	0	closed	closed	closed	0	0	0	23	0	AGflt	A	A	A	A	A	A	A	A	1234	0	Fwd	0	1234	0	0	0	23	0	0		0	0	0	0	0	0	0	0	0	0
t_{90}	0.106843	0	0	0	123	0	0	closed	closed	closed	0	0	0	123	0	AGflt	A	A	A	A	A	A	A	A	1234	0	Fwd	0	1234	0	0	0	123	0	0	1	0	0	0	0	0	0	0	0	0	0
t_{91}	0.108043	0	0	0	123	0	0	closed	closed	closed	0	0	0	123	0	AGflt	A	A	A	A	A	A	A	A	1234	0	Fwd	0	1234	0	0	0	123	0	0	1	0	0	0	0	0	0	0	0	0	A
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
t_{128}	0.152461	0	0	0	123	0	0	closed	closed	closed	0	0	0	123	0	AGflt	A	A	A	A	A	A	A	A	1234	0	Fwd	0	1234	0	0	0	123	0	0	1	0	0	0	0	0	0	0	0	0	A
t_{129}	0.153661	0	0	0	123	0	0	open	closed	closed	0	0	0	123	0	no fault	A	A	A	A	A	A	A	A	1234	0	Fwd	0	1234	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
t_{140}	0.166867	0	0	0	123	0	0	open	closed	closed	0	0	0	123	0	no fault	A	A	A	A	A	A	A	A	1234	0	Fwd	0	1234	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A
t_{141}	0.168067	0	0	0	123	0	0	open	closed	closed	0	0	0	123	0	no fault	A	A	A	A	A	AB	AB	AB	1234	0	Fwd	0	1234	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A
t_{142}	0.169268	0	0	0	123	0	0	open	closed	closed	0	0	0	123	0	no fault	A	A	A	AB	A	AB	AB	AB	1234	0	Fwd	0	1234	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
t_{278}	0.332533	0	0	0	0	0	0	open	closed	closed	0	0	0	0	0	no fault	0	0	0	0	0	0	0	0	0	0	Rev	Fwd	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A
t_{279}	0.333733	0	0	0	0	0	0	open	closed	closed	0	0	0	0	0	no fault	0	0	0	0	0	0	0	0	0	0	Rev	Fwd	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
t_{323}	0.386555	0	0	0	0	0	0	closed	closed	closed	0	0	0	0	0	no fault	0	0	0	0	0	0	0	0	0	0	Fwd	Fwd	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t_{324}	0.387755	0	0	0	0	0	0	closed	closed	closed	0	0	0	0	0	no fault	0	0	0	0	0	0	0	0	0	0	Fwd	Fwd	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Inconsistent decision patterns between these events (i.e., similar condition attributes but the decision attribute *Trip* goes from 0 to A) that can be immediately recognized to have association with tripping signal assertion for CB.

With Visual Basic scripting in the DIAdem environment[25], this decision table has been prepared using a strategy involving such processes as data selection, preprocessing and transformation (construction and discretization) of a simulated raw IEEE-comtrade formatted relay event report which are essential in knowledge discovery in relay database. Some inconsistent patterns are evident in the decision system of which we can recognize to have association with the decision attribute *Trip* going from 0 to A, inferring tripping signal assertion of CB. This shall be one of a number of aspects that shall be considered when protective relay analysis is done using rough set theory technique.

RESULTS AND DISCUSSIONS

When considering merely the relay information system $IS = \langle U, C, V, f \rangle$ instead of the relay decision system $DT = \langle U, C \cup D, V, f \rangle$, the C-equivalence (indiscernibility) relation $U|IND(C)$ is found that comprises the family of all elementary sets of relay events that are indiscernible based on all the condition attributes $c_i \in C$. Table 2 shows the elementary sets with respect to C about U in IS.

When considering the relay decision system $DT = \langle U, C \cup D, V, f \rangle$, the D-equivalence (indiscernibility) relation $U|IND(D)$ is found that comprises the family of all elementary sets of relay events that are indiscernible based on the decision attribute $D = \{Trip\}$.Table 3 shows the elementary sets with respect to D about U in DT.

Discovering Knowledge of Relay Trip Assertion: An important observation pertaining to the the fact that the cardinality of the equivalence class $Trip$ = A, i.e., $card(IND(D)_{Trip\ =\ \mathrm{A}})$, having a non-zero value implies that the distance relay has been subjected to a fault and thus upon which it has acted by sending trip signal to the circuit breaker to open the line and eventually clear the fault.

The cardinality of the equivalence class $Trip$ = A, i.e., $card(IND(D)_{Trip\ =\ \mathrm{A}}) = 188$, indicates the duration the relay is in assertion (set) state for delivering tripping signal to circuit breaker to operate. This corresponds to the relay events that are indiscernible with respect to decision attribute $Trip$ = A and are representative of the continuous time sequence from, as identified by the time codes (events), t_{91} to t_{278}, that is, Relay trip assertion duration,

$$\begin{aligned} T_{trip\ assertion} &= t_{278} - t_{91} \\ &= 0.332533\ \text{s} - 0.108043\ \text{s} \\ &= 0.2245\ \text{s} \end{aligned}$$

This duration is in agreement with the required minimum total time delay the relay takes to reset as specified by the parameter of relay setting $asrt\_trpT$ = 0.18 s in the simulation. Alternatively, using $card(IND(D)_{Trip\ =\ \mathrm{A}}) = 188$, $T_{trip\ assertion}$ can be calculated as,

$$\begin{aligned} T_{trip\ assertion} &= [card(IND(D)_{Trip\ =\ \mathrm{A}}) - 1]^{\mathrm{x}}\ T_{relay\ recording\ time\ step} \\ &= [188 - 1]^{\mathrm{x}}\ 1200\ \text{ms} \\ &= 0.2244\ \text{s} \end{aligned}$$

Table 2: C-equivalence relation about U in $IS = <U, C, V, f>$ for distance protective relay involving zone 1 A-G fault

C-elementary sets (w.r.t. all condition attributes, C), i.e. condition classes of $U \vert IND(C)$		Cardinality, *card* (IND(c))
$\{t_{76}, \dots t_{79}, t_{321}, \dots t_{324}\}$	$IND(C)_1$	8
$\{t_{80}\}$	$IND(C)_2$	1
$\{t_{81}\}$	$IND(C)_3$	1
$\{t_{82}\}$	$IND(C)_4$	1
$\{t_{83}\}$	$IND(C)_5$	1
$\{t_{84}, t_{85}\}$	$IND(C)_6$	2
$\{t_{293}, t_{294}, t_{306}, \dots t_{320}\}$	$IND(C)_7$	17
$\{t_{185}, t_{186}\}$	$IND(C)_8$	2
$\{t_{286}, t_{287}\}$	$IND(C)_9$	2
$\{t_{178}, \dots t_{184}, t_{187}, \dots t_{285}\}$	$IND(C)_{10}$	106
$\{t_{289}\}$	$IND(C)_{11}$	1
$\{t_{290}, t_{291}, t_{292}, t_{295}, \dots t_{298}\}$	$IND(C)_{12}$	7
$\{t_{304}, t_{305}\}$	$IND(C)_{13}$	2
$\{t_{302}\}$	$IND(C)_{14}$	1
$\{t_{299}, t_{300}, t_{301}\}$	$IND(C)_{15}$	3
$\{t_{303}\}$	$IND(C)_{16}$	1
$\{t_{177}\}$	$IND(C)_{17}$	1
$\{t_{176}\}$	$IND(C)_{18}$	1
$\{t_{171}, \dots t_{175}\}$	$IND(C)_{19}$	5
$\{t_{169}, t_{170}\}$	$IND(C)_{20}$	2
$\{t_{168}\}$	$IND(C)_{21}$	1
$\{t_{86}\}$	$IND(C)_{22}$	1
$\{t_{167}\}$	$IND(C)_{23}$	1
$\{t_{88}, t_{89}\}$	$IND(C)_{24}$	2
$\{t_{87}\}$	$IND(C)_{25}$	1
$\{t_{288}\}$	$IND(C)_{26}$	1
$\{t_{166}\}$	$IND(C)_{27}$	1
$\{t_{90}, \dots t_{128}\}$	$IND(C)_{28}$	39
$\{t_{152}\}$	$IND(C)_{29}$	1
$\{t_{129}, t_{130}, \dots t_{140}\}$	$IND(C)_{30}$	12
$\{t_{141}\}$	$IND(C)_{31}$	1
$\{t_{142}, t_{143}\}$	$IND(C)_{32}$	2
$\{t_{144}, t_{145}\}$	$IND(C)_{33}$	2

Table 2: Continue

$\{t_{149}, t_{150}\}$	$IND(C)_{34}$	2
$\{t_{146}, t_{147}, t_{148}\}$	$IND(C)_{35}$	3
$\{t_{151}\}$	$IND(C)_{36}$	1
$\{t_{153}\}$	$IND(C)_{37}$	1
$\{t_{154}, \dots t_{165}\}$	$IND(C)_{38}$	12

Table 3: D-equivalence relation about U in DT = <U, C È D, V, f > of distance protective relay involving zone 1 A-G fault

Equivalence classes (w.r.t. decision attribute, $D = \{Trip\}$), i.e. decision classes of $U \vert IND(D)$		Cardinality, $card(IND(D))$
$\{t_{76}, \dots t_{90}, t_{279}, \dots t_{292}, t_{295}, \dots t_{324}\}$	$IND(D)_{Trip\ =\ 0}$	61
$\{t_{91}, \dots t_{278}\}$	$IND(D)_{Trip\ =\ \mathrm{A}}$	188

Table 4: C_Z-elementary sets $IND(C_Z)$ about U in IS = <U, C, V, f > w.r.t. condition attributes $C_Z \subset C \mid C_Z = \{Zab, Zbc, Zca, Zag, Zbg, Zcg\}$ for distance protective relay exposed to zone 1 A-G fault.

C_Z-elementary sets $IND(C_Z)$ w.r.t. condition attributes, C_Z Ì C ½ $C_Z = \{Zab, Zbc, Zca, Zag, Zbg, Zcg\}$		Cardinality, $card(IND(C))$
$\{t_{76}, \dots t_{85}, t_{168}, \dots t_{324}\}$	$IND(C_Z)_1$	166
$\{t_{86}\}$ $IND(C_Z)_2$	1	
$\{t_{167}\}$ $IND(C_Z)_3$	1	
$\{t_{87}, t_{88}, t_{89}, t_{166}, t_{288}\}$	$IND(C_Z)_4$	5
$\{t_{90}, \dots t_{165}\}$	$IND(C_Z)_5$	76
	249 (U)	

The first relay event in the elementary set of the equivalence class $Trip$ = A, that is t_{91}, tells us the instance where the relay starts to set for triggering the circuit breaker to trip the faulted section of the transmission line. Obviously to make the relay operation more sense, the relay must have picked up the fault prior to event t_{91} and the trip status of the decision attribute $Trip$ should be "0".

Discovering Knowledge of Impedance Element Activation and Fault Characteristics: It is interesting to know whether rough set theory can help in identifying the type and time characteristics of fault to which the relay might have been subjected and the triggered impedance elements of the relay. Since fault behavior is sought after, it is desirable to focus on analyzing the condition attributes relating to faults. Thus, an equivalence relation specifying the C-elementary sets $IND(C)$ of relay events that are indiscernible with respect to impedance-measurement element attributes $C_Z \subset C \mid C_Z = \{Zab, Zbc, Zca, Zag, Zbg, Zcg\}$ ought to be found and as shown in Table 4.

Next, the approximation of the decision system classification is analyzed by calculating in particular the C_Z-lower approximation of classification of D with respect to decision attribute $Trip$ having class "A". The result of overall approximation is shown in Table 5.

From the C_Z-lower approximation of $D_{Trip\ =\ \mathrm{A}}$, it can be seen that the elementary sets of relay events can imply the fault behavior that results in the operation of the distance relay. The first event t_{90} in the sequence of relay events that form the elementary set in the C_Z-lower approximation of $D_{Trip\ =\ \mathrm{A}}$, (i.e. $\{t_{90}, \dots t_{165}\} \cup \{t_{167}\}$) indicates the impedance element activation in response to fault. This is anticipated since the fault initiation leads up to the distance relay's triggering the trip signal during the relay event t_{91} as discussed earlier.

With the impedance element activation corresponding to the relay event t_{90} in mind, we should find what kind of fault that has actually occurred. In order to identify it, we should first of all partition the universe of relay events into C-elementary sets $IND(C)$ with respect to the individual condition attributes pertaining to the various impedance elements as shown in Table 6. Then comparison is made between the C_Z-lower approximation of $D_{Trip=\mathrm{A}}$ and the cz_i-elementary sets $IND(c_{Zi})$ and find in the latter which of the single impedance-element condition attribute $cz_i \in C$ has the relay event t_{90} being an elemental unit (actually the first one) in a particular subset of its discovered elementary sets. This is identified in the Table 6 as an underlined event t_{90} and which the corresponding single attribute $cz_i \in C \vert C = \{Zag\}$ denotes that a phase-a-ground (A-G) fault has occurred. The inferred fault type can be verified by cross checking with the value at the instance of relay event t_{90} of the condition attribute that solely represents the occurance of fault itself, i.e. $c_i \in C|\ c_i = FltType$. In this case $FltType(t_{90})$ = AGflt.

The cardinality $card(IND(cz_i))$ = 76 signifies the end parameters needed to calculate the duration the A-G impedance element takes to last after its activation at relay event t_{90}, i.e.,

$$\begin{aligned} T_{Zag\ activation} &= [card(IND(cz_i)) - 1] \times T_{relay\ recording\ time\ step} \\ &= [76 - 1] \times 1200 \text{ ms} \\ &= 90 \text{ ms} \end{aligned}$$

This A-G fault in as far as distance protection is concerned is more appropriately taken to be related to which zone of protection the relay has acted upon. At event t_{90}, the condition attribute Zag, as apparent from Table 1, gives us the value of "123". By the nature of the distance relay being a quadrilateral type as shown in Fig. 3, this value is considered to be the concurrent occurrence of the A-G fault in all three zones 1, 2 and 3. In quadrilateral distance relay, the encapsulating nature of its operation characteristic suggests that Zag = 123 would mean that zone 1 is the priority indication of fault occurance for which the relay would refer to function correctly according to the preset time for zone 1 operation (usually instantaneously ($Z1onT$) or as constrained by digital sampling of data, i.e. immediately after one time step).

To find the total time duration of the A-G fault, i.e. $T_{fault\ duration}$, by relying on the impedance-element condition attributes as either a set ($C_Z \subset C \mid C_Z = \{Zab, Zbc, Zca, Zag, Zbg, Zcg\}$) or singly ($cz_i \in C$) may not be practical. This is because after the A-G impedance element ($C = \{Zag\}$) triggering at t_{90}, what happen after that is immaterial. The duration of the A-G impedance element activation $T_{Zag\ activation} =$ 90 ms could possibly be only a fraction of $T_{fault\ duration}$ if the cumulative time of the impedance measurement trajectory's entry into and exit from the distance relay's quadrilateral characteristics is within the $T_{fault\ duration}$. The $T_{fault\ duration}$ could also be possibly less than $T_{Zag\ activation}$ if the exit of the impedance measurement trajectory away from the distance relay's quadrilateral characteristics of zone 1 is just a natural progression after the fault has been cleared.

To resolve the above problem, the total time duration of the A-G fault, i.e. $T_{fault\ duration}$, must take into account the the condition attribute that solely represents the occurance of fault itself, i.e. $c_i \in C \mid c_i = \{FltType\}$. We can do this by doing *self-approximation* of $c_i = \{FltType\}$. In other words, the equivalence relation of the $c_i = \{FltType\}$ is approximated with respect to itself using the C-elementary sets $IND(C)$ of relay events that are indiscernible with respect to $c_i \in C|\ c_i = \{FltType\}$. The purpose of self-approximation is simply to extract information about the 'crisp' $C_{FltType}$-lower approximation of $C_{FltType=\text{AGflt}}$ so that the information can be directly used for analysis especially in automation. The equivalence relation specifying c_i-elementary sets $IND(c_i)$ with respect to own-condition attribute, $c_i \in C|\ c_i = \{FltType\}$ and the self-approximation analysis of $c_{i,FltType\ =\ \text{AGflt}}$ are shown shown in Tables 7 and 8 respectively. Clearly, as expected, when the upper- and lower-approximations are equal (i.e., boundary region empty), the self-target set $c_i = \{FltType\}$ is perfectly approximated and said to be accurately *definable* in U with an accuracy

$$\begin{aligned} &\alpha_{c_i}(c_i) \\ &= \left|\underline{c}_i\, c_{i,FltType=\text{AGflt}}\right| / \left|\overline{c}_i\, c_{i,FltType=\text{AGflt}}\right| \\ &= card(\underline{c}_i\, c_{i,FltType=\text{AGflt}}) / card(\overline{c}_i\, c_{i,FltType=\text{AGflt}}) \\ &= 47/47 = 1 \end{aligned}$$

(not quite a surprise!). In Table 8, it is evident that information in the lower-approximation can determine the fault characteristics needed. The first event t_{82} in the lower-approximation set indicates the identifier for the A-G fault inception. Refering to Table 1, t_{82} points to the instance 0.097239 s at which the A-G fault starts to pronounce. The expected fault duration $T_{fault\ duration}$ can be easily calculated as follows:

$$\begin{aligned} T_{fault\ duration} &= [card(\underline{c}_i\, c_{i,FltType=\text{AGflt}}) - 1] \times T_{relay\ recording\ time\ step} \\ &= [47 - 1] \times 1200 \text{ ms} \\ &= 55.2 \text{ ms} \end{aligned}$$

Or similarly,

$$\begin{aligned} T_{fault\ duration} &= t_{128} - t_{82} \text{ (the end time codes taken from } \underline{c}_i\, c_{i,FltType=\text{AGflt}}) \\ &= 0.152461\text{s} - 0.097239 \text{ s} \\ &= 55.2 \text{ ms} \end{aligned}$$

In Table 6, the other single impedance-element attributes ($cz_i = \{Zab\}$, $cz_i = \{Zbc\}$, $cz_i = \{Zcd\}$, $cz_i = \{Zbg\}$ and $cz_i = \{Zcg\}$) have their elementary sets similarly representing some particularly undetected faults of the types corresponding to the attributes. The fact that $card(IND(cz_i))$ = 249 measures exactly similar size as the entire universe U of relay events indicates that no fault of the corresponding types have been imminent. In fact $card(IND(cz_i))$ = 249 measures the indiscernibility of the entire relay events with respect to the value of "0" in every single impedance-element attribute except $cz_i = \{Zag\}$.

Summary: In this paper rough-set-based data mining strategy in KDD has been successfully formulated to intelligently discover the knowledge of:

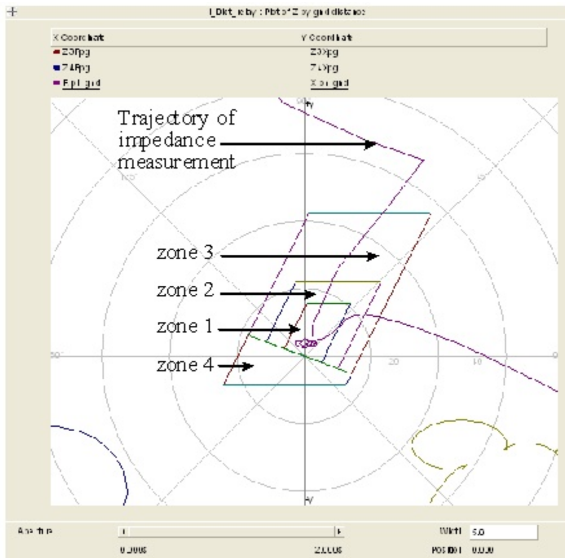

Fig. 3: Distance protective relay operation characteristic with impedance measurement trajectory

Table 5: C_Z-approximations of classification of D with respect to decision attribute $Trip$ = A and condition attributes $C_Z \in C \mid C_Z = \{Zab, Zbc, Zca, Zag, Zbg, Zcg\}$

	C_Z-upper approximation of $D_{Trip\,=\,A}$, i.e. ($\overline{C_Z}\, D_{Trip\,=\,A}$)	C_Z-lower approximation of $D_{Trip\,=\,A}$, i.e. ($\underline{C_Z}\, D_{Trip\,=\,A}$)	C_Z-boundary of $D_{Trip\,=\,A}$, i.e. $BNc_Z(D_{Trip\,=\,A})$	Outside
	$\{t_{76}, \dots t_{85}, t_{168}, \dots t_{324}\} \cup \{t_{167}\} \cup \{t_{87}, t_{88}, t_{89}, t_{166}, t_{288}\} \cup \{t_{90}, \dots t_{165}\}$	$\{t_{90}, \dots t_{165}\} \cup \{t_{167}\}$	$\{t_{76}, \dots t_{85}, t_{168}, \dots t_{324}\} \cup \{t_{87}, t_{88}, t_{89}, t_{166}, t_{288}\}$	$\{t_{86}\}$
Cardinality	248	77	171	1
		249 (Cardinality of U)		

Table 6: c_{Zi}-elementary sets $IND(c_{Zi})$ with respect to the individual condition attributes.

w.r.t. single fault-type condition attribute, $cz_i \in C$	cz_i-elementary sets $IND(cz_i)$	Cardinality, $card(IND(cz_i))$
$C = \{Zab\}$	$\{t_{76}, \dots t_{324}\}$	249
$C = \{Zbc\}$	$\{t_{76}, \dots t_{324}\}$	249
$C = \{Zcd\}$	$\{t_{76}, \dots t_{324}\}$	249
$C = \{Zag\}$	$\{t_{76}, \dots t_{85}, t_{168}, \dots t_{324}\}$ $\{t_{86}, t_{167}\}$ $\{t_{87}, t_{88}, t_{89}, t_{166}, t_{288}\}$ $\{t_{90} \dots t_{165}\}$	166 2 5 76
$C = \{Zbg\}$	$\{t_{76}, \dots t_{324}\}$	249
$C = \{Zcg\}$	$\{t_{76}, \dots t_{324}\}$ $\{t_{167}\}$	248 1

Table 7: c_i-elementary sets $IND(c_i)$ about U in $IS = <U, C, V, f>$ w.r.t. condition attributes $c_i \in C \mid c_i = \{FltType\}$ for distance protective relay exposed to zone 1 A-G fault

c_i-elementary sets $IND(c_i)$ w.r.t. condition attributes, $c_i \in C \mid c_i = \{FltType\}$		Cardinality, $card(IND(c_i))$
$\{t_{76}, \dots t_{81}, t_{129}, \dots t_{324}\}$	$IND(c_i)_1$	202
$\{t_{82}, \dots t_{128}\}$	$IND(c_i)_2$	47
		249 (U)

Table 8: c_i-self-approximations of classification of c_i with respect to own-condition attribute $c_i \in C|\ c_i = \{FltType\}$ having value $FltType$ = AGflt

	c_i-upper approximation of $c_{\_\ ype = \text{AGflt}}$, i.e. ($c_i$ $c_{i,FltType = \text{Agflt}}$)	c_i-lower approximation of $c_{\ ype = \text{AGflt}}$, i.e. ($c_{i\ FltType = \text{Agflt}}$)	c_i-boundary of $c_{i,FltType = \text{AGflt}}$, i.e. $BN_{ci}(c_{i,FltType = \text{Agflt}})$	Outside
	$\{t_{82}, \ldots\ t_{128}\}$	$\{t_{82}, \ldots\ t_{128}\}$	Ø	$\{t_{76}, \ldots\ t_{324}\}$
Cardinality	47	47	0	202
		249 (Cardinality of U)		

- distance relay trip assertion,
- impedance element activation,
- fault characteristics,

Using rough set theory, the uncertainty and vagueness in the relay event report can be resolved using the concepts of discernibility, elementary sets and set approximations. The combination of rough-set-based data mining and event report resident in digital protective relay has made it possible to lessen the burden born by protection engineers having to deal with very complex implementations of protection system analysis due to massive quantities of data coming from diverse types of intelligent electronic devices.

Unlike some artificial intelligence aproaches like artificial nueral network and decision tree in which the data mining analysis is "population-based" and single since it is common to the entire population of training data set, the rough set approach adopts an individually-event-based" paradigm in which detailed time tracking analysis of relay operation has been successfully performed.

REFERENCES

1. Kezunovic, M., 2001. Section II: Equipment characteristics. IEEE tutorial on automated fault analysis, Texas A&M University, College Station, USA, 5-9.
2. Makki, A. and M.R. Makki, 2001. Transient Data Repository. Fault and Disturbance Analysis Conference, April 29-30, Atlanta, Georgia, USA, 1-7.
3. Negnevitsky, M. and V. Pavlovsky, 2005. Neural networks approach to online identification of multiple failures of protection systems. IEEE Trans. on Power Delivery, 20(2): 588-594.
4. Zhang, N. and M. Kezunovic, 2006. A real time fault analysis tool for monitoring operation of transmission line protective relay. ELSEVIER Electric Power Systems Research, 77(3-4): 361-370.
5. MacArthur, S.D.J., J.R. McDonald, S.C. Bell and G.M. Burt, 1995. Expert systems and model based reasoning for protection performance analysis. IEE Colloquium on Artificial Intelligence Applications in Power Systems, 20: 1/1- ¼.
6. Bell, S.C., S.D.J. McArthur, J.R. McDonald, G.M. Burt, R. Mather and T. Cumming, 1998. Model-based analysis of protection system performance. IEE Proc. Generation, Transmission and Distribution, 145(5): 547 – 552.
7. Davidson, E.M., S.D.J. McArthur, J.R. McDonald, 2003. A toolset for applying model-based reasoning techniques to diagnostics for power systems protection. IEEE Trans. on Power Systems, 18(2): 680 – 687.
8. Dumitrescu, M., T. Munteanu, A.P. Ulmeanu, 2004. Electric-power protection system fuzzy critical analysis. IEEE Proc. 2nd International Conference on Intelligent Systems, 22-24 June, 1: 322-325.
9. Kezunovic, M. and X. Luo, 2005. Automated analysis of protective relay data. CIRED 18th International Conference on Electricity Distribution, Session 3, June 6-9, Turin.
10. Luo, X. and M. Kezunovic, 2005. Automated Analysis of Digital Relay Data Based on Expert System. IEEE.
11. Zhang, N. and M. Kezunovic, 2004. Verifying the protection system operation using an advanced fault analylsis tool combined with the event tree analysis. Northern American Power Symposium NAPS, Moscow.
12. Behrendt, K.C. and M.J. Dood, 200X. Substation relay data and communication. Technical paper, Schweitzer Engineering Laboratories Inc., Pullman, Washington, USA. Available: http:/www.selinc.com/techpprs.htm.
13. Smith, L., 2001. Section I: Requirement, utility perspective. IEEE tutorial on automated fault analysis, Texas A&M University, College Station, USA, 1-4.
14. Pawlak, Z., 2002. Rough set and intelligent data analysis. Elsevier International Journal of Information Sciences, 147: 1-12.
15. Hor, C.L., P.A. Crossley and S.J. Watson, 2007. Building knowledge for substation-based decision support using rough sets. IEEE Trans. Power Delivery, 22(3): 1372-1379.
16. Xu, X. and J.F. Peters, 2002. Rough set methods in power system fault classification. IEEE Proc. Canadian Conference on Electrical and Computer Engineering. 100-105.

17. Magro, M.C. and P. Pinccti, 2004. Diagnosis for intelligent field devices using the rough set theory. IEEE Proc. Instumentation and Measurement Technology Conference, IMTC 2004, May 18-20, Como, Italy, 2060-2063.
18. Pawlak, Z., 1982. Rough sets. International Journal of Computer and Information Sciences, 11(5): 341-356.
19. Walczak, B. and D.L. Massart, 1999. Rough sets theory: tutorial. Chemometrics and Intelligent Laboratory Systems, 47(1): 1-16.
20. Zhai, L.Y., L.P. Khoo and S.C. Fok, 2006. Knowledge acquisition and uncertainty in fault diagnosis: A rough sets perspective, in Triantaphyllor, E, and Felici, G. (Eds.): Data mining and knowledge discovery approaches based on rule induction techniques, Massive Computing Series, Springer, Heidelberg,Germany, 359-394.
21. Drwal, G. and M. Sikora, 2004. Fuzzy Decision Support Systemwith Rough Set Based Rules Generation Method. In: Tsumoto, S., et al. (editors.) RSCTC (2004), LNAI 3066, 727–732. Springer-Verlag Berlin Heidelberg.
22. Freitas, A.A., 2002. Data Mining and Knowledge Discovery with Evolutionary Algorithms. Springer-Verlag Berlin Heidelberg New York, 1.
23. Kusiak, A., 2001. Rough set theory: A data mining tool for semiconductor manufacturing. IEEE Transactions on Electronics Packaging Manufacturing, 24(1): 44-50.
24. PSCAD., 2003. PSCAD User's Guide. Manitoba, Manitoba HVDC Research Centre Inc. 244 Cree Cresent, Winnipeg, Manitoba, Canada.
25. National Instruments. 2005. DIAdem: Data mining, analysis, and report generation. National Instruments Corporation, Austin, Texas, USA. Doc. No. 373082D-01.