Novel Rule Base Development from IED-Resident Big Data for Protective Relay Analysis Expert System

RESEARCH-ARTICLE

Mohammad Lutfi Othman[1*], Ishak Aris[1] and Thammaiah Ananthapadmanabha[2]
Show details

Abstract

Many Expert Systems for intelligent electronic device (IED) performance analyses such as those for protective relays have been developed to ascertain operations, maximize availability, and subsequently minimize misoperation risks. However, manual handling of overwhelming volume of relay resident big data and heavy dependence on the protection experts' contrasting knowledge and inundating relay manuals have hindered the maintenance of the Expert Systems. Thus, the objective of this chapter is to study the design of an Expert System called Protective Relay Analysis System (PRAY), which is imbedded with a rule base construction module. This module is to provide the facility of intelligently maintaining the knowledge base of PRAY through the prior discovery of relay operations (association) rules from a novel integrated data mining approach of Rough-Set-Genetic-Algorithm-based rule discovery and Rule Quality Measure. The developed PRAY runs its relay analysis by, first, validating whether a protective relay under test operates correctly as expected by way of comparison between hypothesized and actual relay behavior. In the case of relay maloperations or misoperations, it diagnoses presented symptoms by identifying their causes. This study illustrates how, with the prior hybrid-data-mining-based knowledge base maintenance of an Expert System, regular and rigorous analyses of protective relay performances carried out by power utility entities can be conveniently achieved.

Keywords: association rule, data mining, digital protective relay, expert system, power system protection analysis, rough set theory

1. Introduction

According to the IEEE Working Group D10 of the Line Protection Subcommittee, Power System Relaying Committee, Expert Systems have been proposed since early 1980s to be potential tools for engineers to develop intelligent performance analysis systems for the intelligent electronic devices (IEDs) such as protective relays [1]. Some of the works where protection performance analyses can be identified are in the area of offline tasks such as settings coordination, postfault analysis, and fault diagnosis [2–13].

Kezunovic et al. [6] explain the substation automated fault analysis using Expert System method based on the retrieved disturbance data acquired by digital fault recorders (DFRs). This fault analysis helps protection engineers identify the correctness of protective relay operation. **Figure 1** illustrates the block diagram of the Expert System. The knowledge base in the CLIPS (an Expert System shell) rules used in the forward chaining inference engine using processed data is built by interviewing experts, using an empirical approach based on Electromagnetic Transient Program (EMTP) simulation and utilizing actual big field substation data.

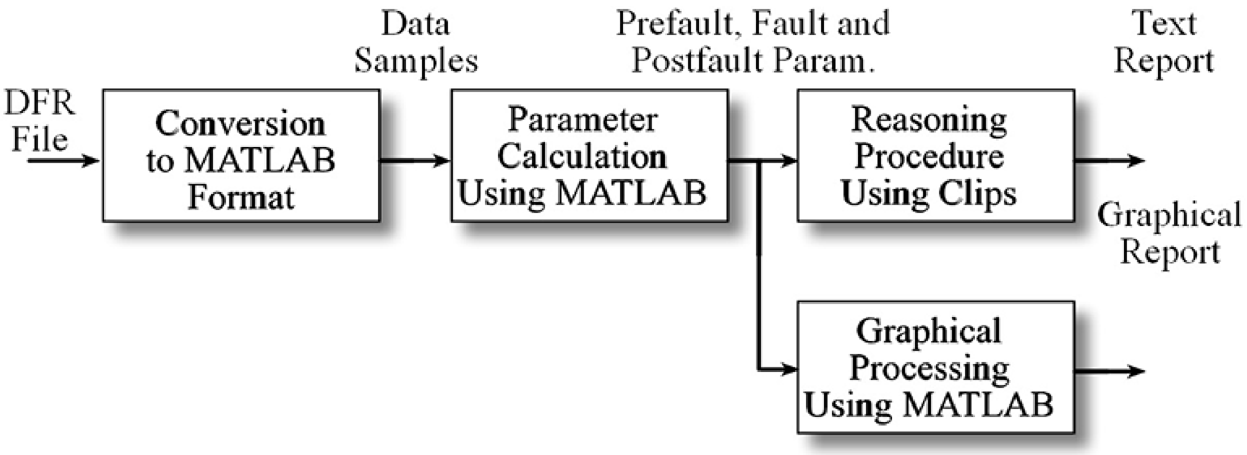

Data Samples
Prefault, Fault and Postfault Param.
Text Report
DFR File
Conversion to MATLAB Format
Parameter Calculation Using MATLAB
Reasoning Procedure Using Clips
Graphical Report
Graphical Processing Using MATLAB

FIGURE 1.

The Expert System block diagram [6].

Luo and Kezunovic's [10] implementation of the Expert System in automated protection analysis is more specifically tailored at detailed analysis of a specific protective relay by relying on recorded big data found only within it. **Figure 2** illustrates the block diagram of the analysis system created based on CLIPS language within Visual C++ framework. The analysis system is developed revolving around the strategy of comparing predicted (hypothesized) and actual (factual) protection operation in terms of statuses and corresponding timings of logic operands. Any matching between the predicted and actual protection operations validates the correctness of the actual status and timing of that operand. Otherwise, certain misoperation is identified, and diagnosis is initiated to trace the reasons. Predicted statuses and timings of active logic operands are basically a hypothesization of relay operations, which is done by way of forward chaining reasoning. They form the knowledge base in the rules used in the CLIPS inference engine.

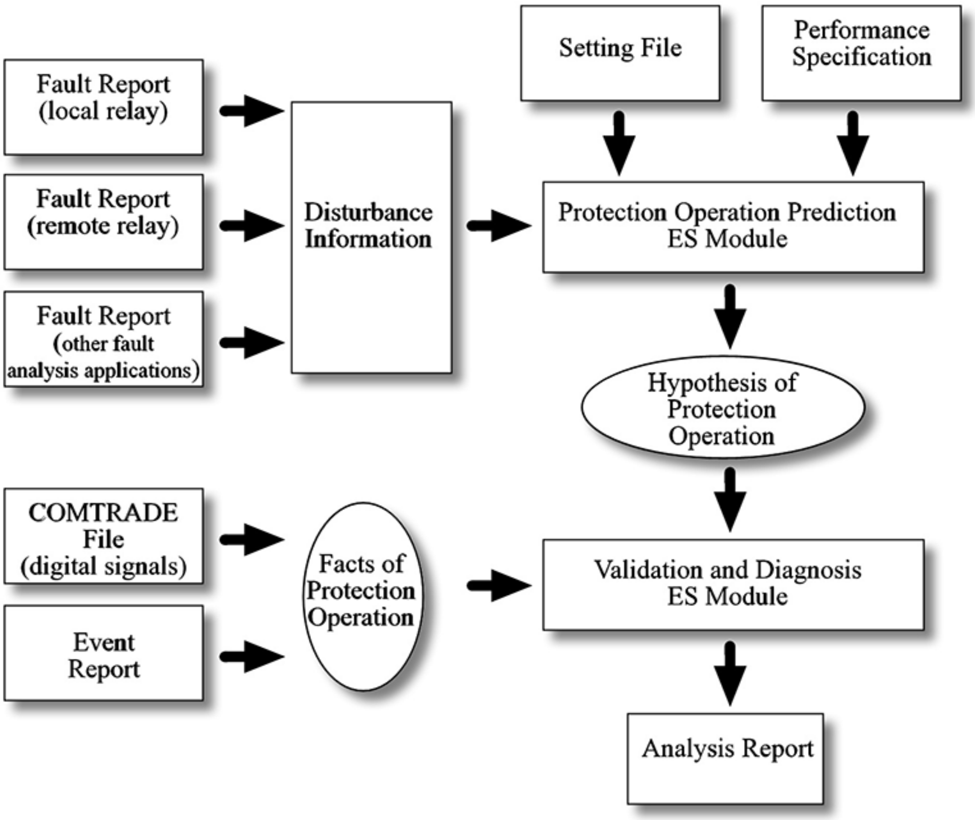

Setting File
Performance Specification
Fault Report (local relay)
Fault Report (remote relay)
Fault Report (other fault analysis applications)
Disturbance Information
Protection Operation Prediction ES Module
Hypothesis of Protection Operation
COMTRADE File (digital signals)
Event Report
Facts of Protection Operation
Validation and Diagnosis ES Module
Analysis Report

FIGURE 2.

The Expert System block diagram for validation and diagnosis of protective relay [10].

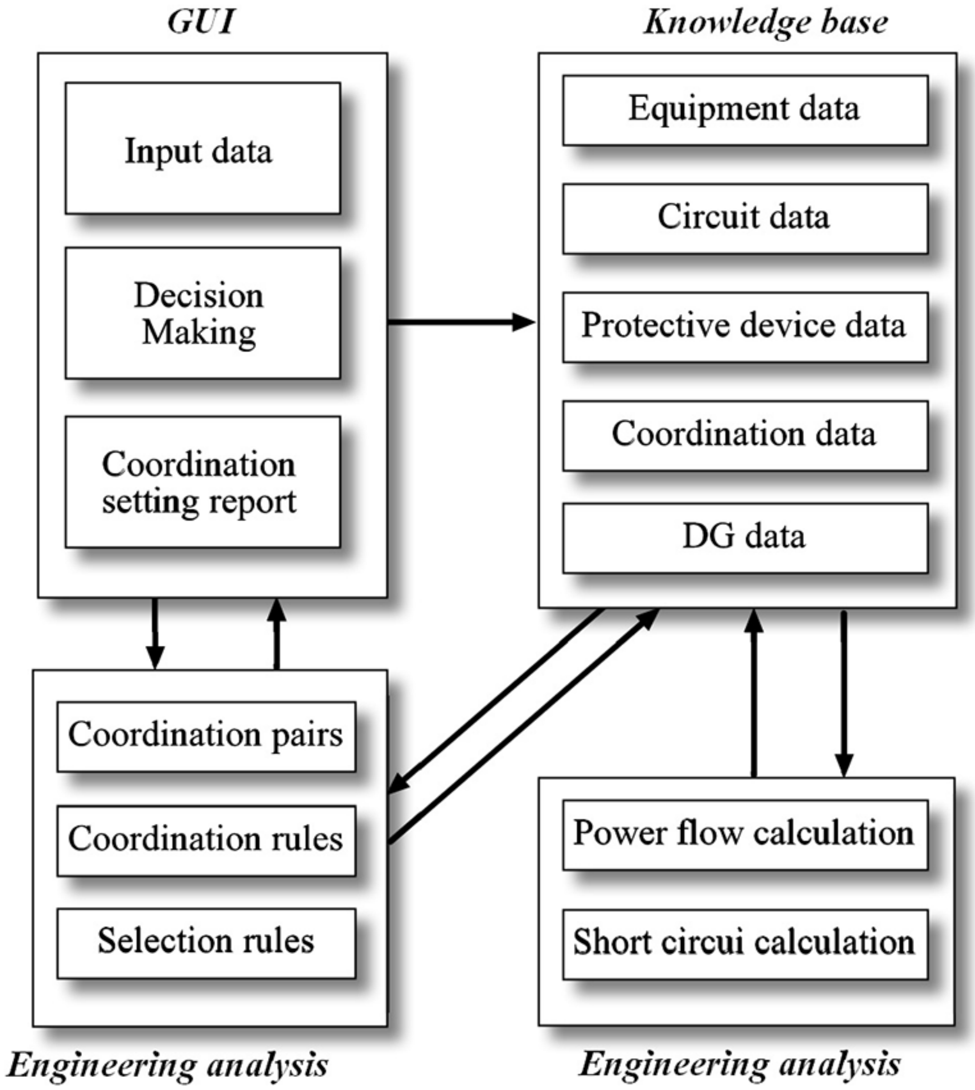


FIGURE 3.

Structure of Expert System for protection coordination [13].

Tuitemwong and Premrudeepreechacharn [13] implement ES analysis for improving protection coordination settings of protective devices in distribution system under the presence of distributed generators (DG). By way of selecting suitable protection coordination settings, this analysis system determines the correct protection system performance in a DG-present power distribution system. The proposed structure of ES is shown in **Figure 3**. The inference engine uses coordination rules and selection rules to generate satisfactory coordination settings based on the processed equipment data, circuit data, protection data, and DG data in the knowledge base. In the case of conflicting settings, the user can make his own decision. The rules are set for the specific distribution system protection and maybe changed when necessary.

The common problem with the aforementioned implementation of rule-based Expert System in protection system analysis is the difficult upgrading of its knowledge base that is made up of “if-then” rules used for decision-making inference engine. Upgrading by expansion and refinement are necessary so as to adapt the Expert System to the continuously changing power network topologies, protection strategies, and multiplicity in protective relay functions [14]. However, acquiring knowledge of relay operation characteristics for upgrading of the knowledge base has not been an easy task due to

i. the burdensome manual handling of voluminous protective relay stored data and

ii. the heavy dependence on the protection experts’ differing knowledge and inundating relay manuals.

It is beneficial if a novel technique could be formulated so as to relieve the untoward effort needed to acquire knowledge in building and maintaining the knowledge base. This technique should allow adjustment of knowledge base by training a protective relay device for as many disturbances as exhaustively possible in order to produce a complete inventory of rules. To help realize this, the authors’ previous work of an integrated data mining approach under the Knowledge Discovery in Database (KDD) framework shall be the prior step before the eventual Expert System knowledge base upgrading strategy is subsequently performed [15–17].

2. Integrated data mining approach to hypothesize expected relay behavior from recorded relay event report

Under the KDD framework, Othman et al. [15–17] investigate the implementation of a novel integrated data mining approach under supervised learning in order to discover the knowledge (or “hypothesize”) and the expected relay behavior. This knowledge extraction from the resident large event reports of a digital distance protective relay comes in the form of association rules as shown in**Figure 4**. The integrated data mining encompasses the adoption of the following computational intelligence methods:

i. Rough set theory: Used to *select* the minimal subsets (i.e., reduction) of attributes while maintaining the original syntax of the relay’s big data of event report.

ii. Genetic algorithm: Used to *explore* the optimal sets of the above subsets of reduced attributes from which simple yet accurate prediction rules (i.e., decision algorithm) can be constructed.

iii. Rule quality measure: Used to *extract* the pertinent association rule from a host of the above original population of prediction rules to determine tripping logic of relay upon fault detection. This is what is referred as hypothesization of protective relay operation. This final version of knowledge representation shall be the main constituent for the Expert System knowledge base.

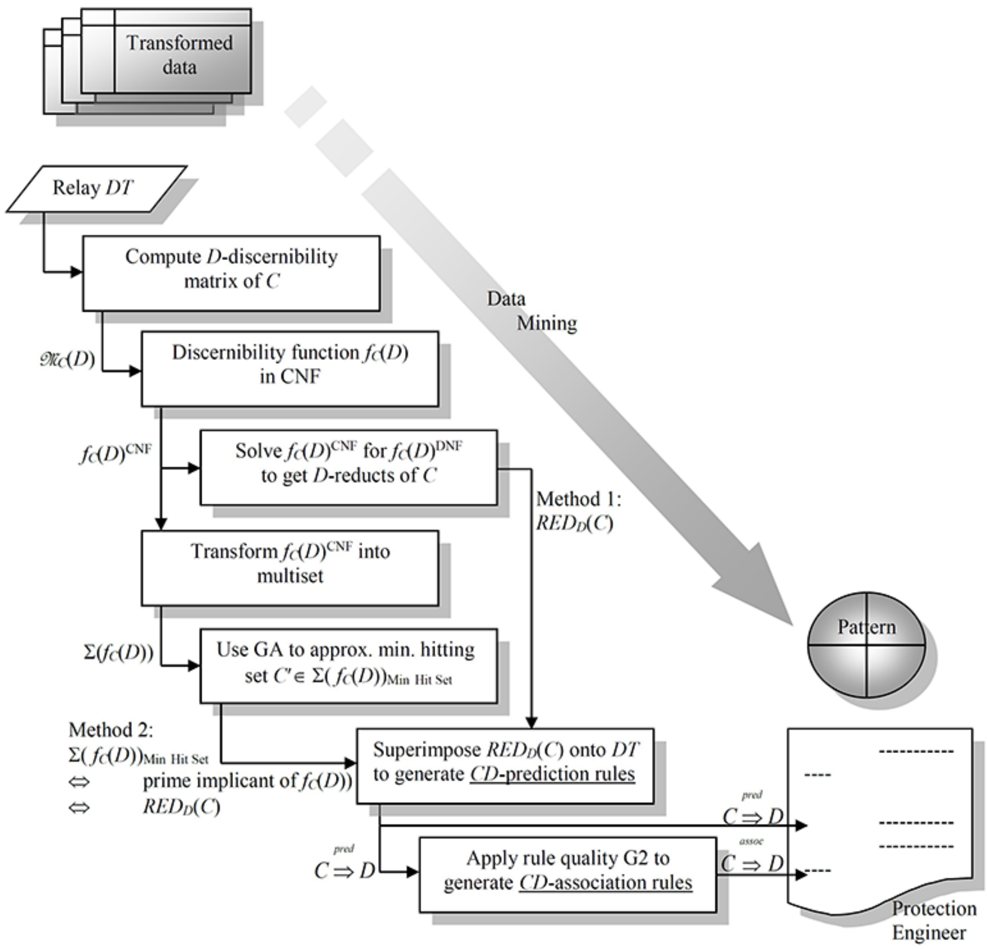
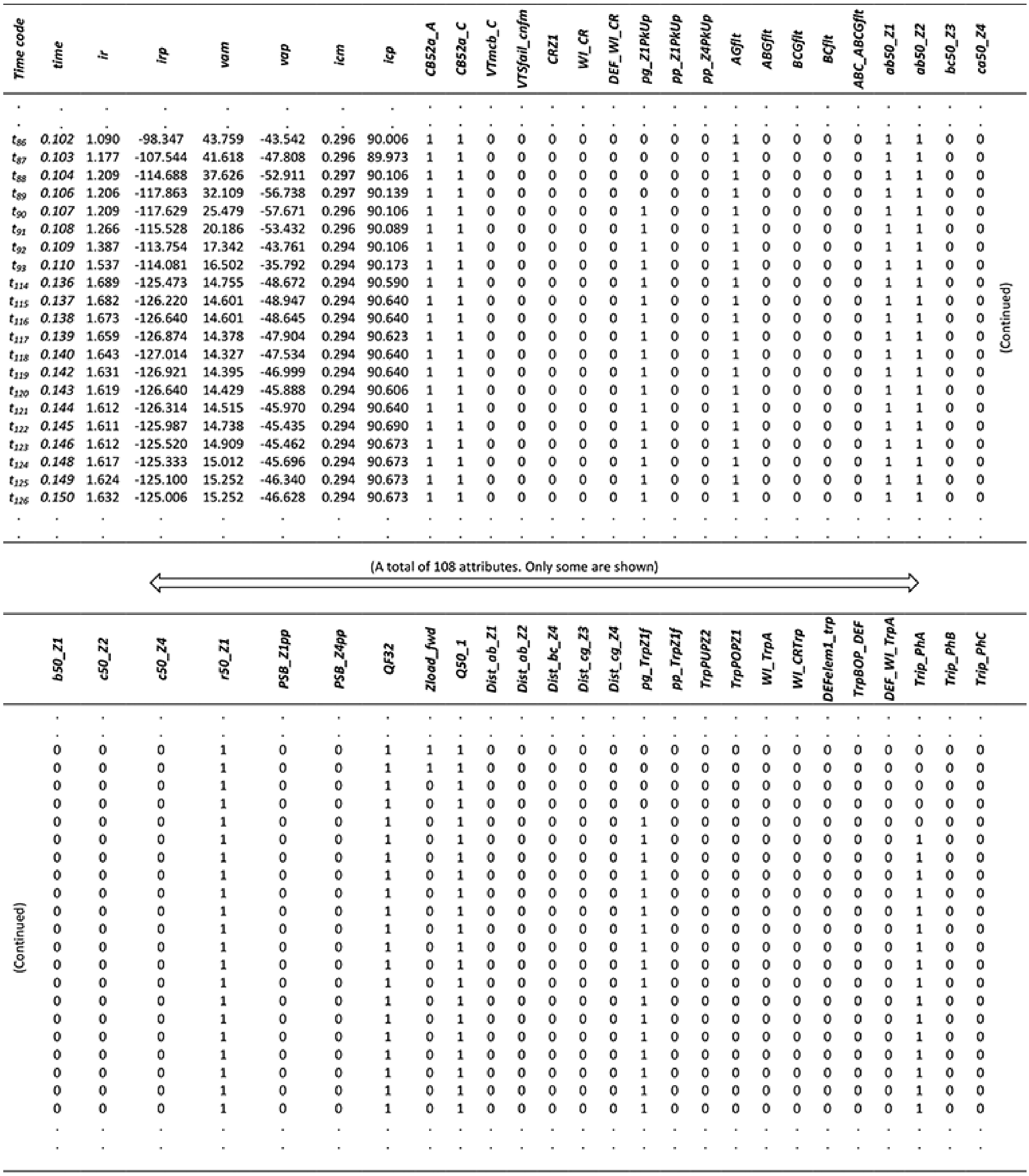

Transformed data
Relay DT
Compute D-discernibility matrix of C
Data Mining
$\mathcal{M}_C(D)$
Discernibility function $f_C(D)$ in CNF
$f_C(D)^{\text{CNF}}$
Solve $f_C(D)^{\text{CNF}}$ for $f_C(D)^{\text{DNF}}$ to get D-reducts of C
Method 1: $RED_D(C)$
Transform $f_C(D)^{\text{CNF}}$ into multiset
$\Sigma(f_C(D))$
Use GA to approx. min. hitting set $C' \in \Sigma(f_C(D))_{\text{Min Hit Set}}$
Pattern
Method 2: $\Sigma(f_C(D))_{\text{Min Hit Set}}$
$\Leftrightarrow$ prime implicant of $f_C(D)$
$\Leftrightarrow$ $RED_D(C)$
Superimpose $RED_D(C)$ onto DT to generate CD-prediction rules
$C \overset{pred}{\Rightarrow} D$
Apply rule quality G2 to generate CD-association rules
$C \overset{pred}{\Rightarrow} D$
$C \overset{assoc}{\Rightarrow} D$
Protection Engineer

FIGURE 4.

Data mining analysis steps in hypothesizing distance relay operation characteristics from big relay event data.

In the study, the large event report is a PSCAD-simulated raw operation recording of an AREVA-modeled distance protective relay as shown in **Table 1** (only a portion of time events is shown to reduce page usage). This big data, which is prior to data preparation, is a representation of the relay's decision system (*DS*) for zone 1 A–G fault—the so-called predata-preparation *DS* [18].



Time code	*time*	*ir*	*irp*	*vam*	*vap*	*icm*	*icp*	*CB52a_A*	*CB52a_C*	*VTmcb_C*	*VTSfail_cnfm*	*CRZ1*	*WI_CR*	*DEF_WI_CR*	*pg_Z1PkUp*	*pp_Z1PkUp*	*pp_Z4PkUp*	*AGflt*	*ABGflt*	*BCGflt*	*BCflt*	*ABC_ABCGflt*	*ab50_Z1*	*ab50_Z2*	*bc50_Z3*	*ca50_Z4*
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
t_{86}	*0.102*	1.090	-98.347	43.759	-43.542	0.296	90.006	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0
t_{87}	*0.103*	1.177	-107.544	41.618	-47.808	0.296	89.973	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0
t_{88}	*0.104*	1.209	-114.688	37.626	-52.911	0.297	90.106	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0
t_{89}	*0.106*	1.206	-117.863	32.109	-56.738	0.297	90.139	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0
t_{90}	*0.107*	1.209	-117.629	25.479	-57.671	0.296	90.106	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0
t_{91}	*0.108*	1.266	-115.528	20.186	-53.432	0.296	90.089	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0
t_{92}	*0.109*	1.387	-113.754	17.342	-43.761	0.294	90.106	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0
t_{93}	*0.110*	1.537	-114.081	16.502	-35.792	0.294	90.173	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0
t_{114}	*0.136*	1.689	-125.473	14.755	-48.672	0.294	90.590	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0
t_{115}	*0.137*	1.682	-126.220	14.601	-48.947	0.294	90.640	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0
t_{116}	*0.138*	1.673	-126.640	14.601	-48.645	0.294	90.640	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0
t_{117}	*0.139*	1.659	-126.874	14.378	-47.904	0.294	90.623	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0
t_{118}	*0.140*	1.643	-127.014	14.327	-47.534	0.294	90.640	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0
t_{119}	*0.142*	1.631	-126.921	14.395	-46.999	0.294	90.640	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0
t_{120}	*0.143*	1.619	-126.640	14.429	-45.888	0.294	90.606	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0
t_{121}	*0.144*	1.612	-126.314	14.515	-45.970	0.294	90.640	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0
t_{122}	*0.145*	1.611	-125.987	14.738	-45.435	0.294	90.690	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0
t_{123}	*0.146*	1.612	-125.520	14.909	-45.462	0.294	90.673	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0
t_{124}	*0.148*	1.617	-125.333	15.012	-45.696	0.294	90.673	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0
t_{125}	*0.149*	1.624	-125.100	15.252	-46.340	0.294	90.673	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0
t_{126}	*0.150*	1.632	-125.006	15.252	-46.628	0.294	90.673	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

(Continued)

(A total of 108 attributes. Only some are shown)

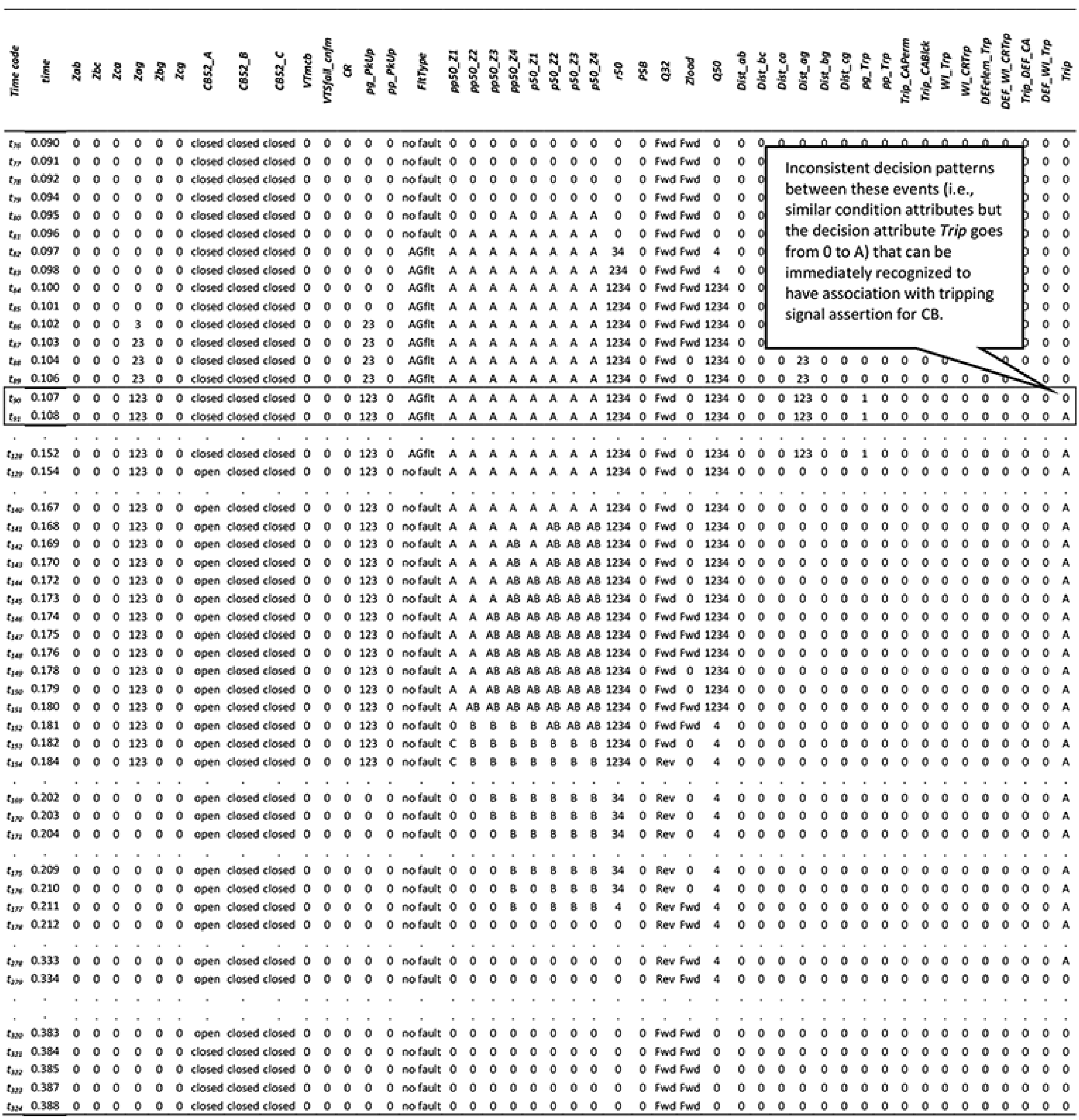
(Continued)

b50_Z1	*c50_Z2*	*c50_Z4*	*r50_Z1*	*PSB_Z1pp*	*PSB_Z4pp*	*QF32*	*Zload_fwd*	*Q50_1*	*Dist_ab_Z1*	*Dist_ab_Z2*	*Dist_bc_Z4*	*Dist_cg_Z3*	*Dist_cg_Z4*	*pg_TrpZ1f*	*pp_TrpZ1f*	*TrpPUPZ2*	*TrpPOPZ1*	*WI_TrpA*	*WI_CRTrp*	*DEFelem1_trp*	*TrpBOP_DEF*	*DEF_WI_TrpA*	*Trip_PhA*	*Trip_PhB*	*Trip_PhC*
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0	0	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

TABLE 1.

Predata-preparation of distance protective relay's decision system for zone 1 A-G fault (only a portion of attribute columns (from a total of 108) and time events are shown to reduce page usage).

The decision system is an information table of event report that can be considered as a pair of finite and nonempty set (U, A). U is the universe of objects (i.e., time tagged relay events t_n, thus called event report) and A is the set of attributes {e.g., *ir, irp, vam, iam, ibm, icm, CB52a_B, CB52b_B, VTmcb_B, CRZ4, pg_Z3PkUp, pg_Z4PkUp, pp_Z1PkUp, pp_Z2PkUp, AGflt, c50_Z1, b50_Z3, Dist_ab_Z2, pg_TrpZ1f, TrpBOPZ1, WI_CRTrp, Trip_PhA*, etc.}. Each attribute $a \in A$ defines an information function such that, f_a: $U \rightarrow V_a$, where V_a is the set of values of the attribute a, called the domain of a. For instance, the set of values of the attribute *pg_Z1PkUp* (the "zone 1 ground distance pick-up" element) is expressed as *pg_Z1PkUp: U* → {0, 1}, which defines the relay element's active states according to the presence of ground fault in the protected section of transmission line (i.e., no-fault present or zone-1-ground-fault present).



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}	0.090	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}	0.091	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
t_{78}	0.092	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
t_{79}	0.094	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
t_{80}	0.095	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
t_{81}	0.096	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
t_{82}	0.097	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
t_{83}	0.098	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
t_{84}	0.100	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
t_{85}	0.101	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
t_{86}	0.102	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	0	0
t_{87}	0.103	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	0	0
t_{88}	0.104	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}	0.106	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	0
t_{90}	0.107	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}	0.108	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}	0.152	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}	0.154	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}	0.167	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}	0.168	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}	0.169	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_{143}	0.170	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_{144}	0.172	0	0	0	123	0	0	open	closed	closed	0	0	0	123	0	no fault	A	A	A	AB	AB	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_{145}	0.173	0	0	0	123	0	0	open	closed	closed	0	0	0	123	0	no fault	A	A	A	AB	AB	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_{146}	0.174	0	0	0	123	0	0	open	closed	closed	0	0	0	123	0	no fault	A	A	AB	AB	AB	AB	AB	AB	1234	0	Fwd	Fwd	1234	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A
t_{147}	0.175	0	0	0	123	0	0	open	closed	closed	0	0	0	123	0	no fault	A	A	AB	AB	AB	AB	AB	AB	1234	0	Fwd	Fwd	1234	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A
t_{148}	0.176	0	0	0	123	0	0	open	closed	closed	0	0	0	123	0	no fault	A	A	AB	AB	AB	AB	AB	AB	1234	0	Fwd	Fwd	1234	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A
t_{149}	0.178	0	0	0	123	0	0	open	closed	closed	0	0	0	123	0	no fault	A	A	AB	AB	AB	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_{150}	0.179	0	0	0	123	0	0	open	closed	closed	0	0	0	123	0	no fault	A	A	AB	AB	AB	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_{151}	0.180	0	0	0	123	0	0	open	closed	closed	0	0	0	123	0	no fault	A	AB	AB	AB	AB	AB	AB	AB	1234	0	Fwd	Fwd	1234	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A
t_{152}	0.181	0	0	0	123	0	0	open	closed	closed	0	0	0	123	0	no fault	0	B	B	B	B	AB	AB	AB	1234	0	Fwd	Fwd	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A
t_{153}	0.182	0	0	0	123	0	0	open	closed	closed	0	0	0	123	0	no fault	C	B	B	B	B	B	B	B	1234	0	Fwd	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A
t_{154}	0.184	0	0	0	123	0	0	open	closed	closed	0	0	0	123	0	no fault	C	B	B	B	B	B	B	B	1234	0	Rev	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
t_{169}	0.202	0	0	0	0	0	0	open	closed	closed	0	0	0	0	0	no fault	0	0	B	B	B	B	B	B	34	0	Rev	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A
t_{170}	0.203	0	0	0	0	0	0	open	closed	closed	0	0	0	0	0	no fault	0	0	B	B	B	B	B	B	34	0	Rev	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A
t_{171}	0.204	0	0	0	0	0	0	open	closed	closed	0	0	0	0	0	no fault	0	0	0	B	B	B	B	B	34	0	Rev	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
t_{175}	0.209	0	0	0	0	0	0	open	closed	closed	0	0	0	0	0	no fault	0	0	0	B	B	B	B	B	34	0	Rev	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A
t_{176}	0.210	0	0	0	0	0	0	open	closed	closed	0	0	0	0	0	no fault	0	0	0	B	0	B	B	B	34	0	Rev	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A
t_{177}	0.211	0	0	0	0	0	0	open	closed	closed	0	0	0	0	0	no fault	0	0	0	B	0	B	B	B	4	0	Rev	Fwd	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A
t_{178}	0.212	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_{278}	0.333	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}	0.334	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_{320}	0.383	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	Fwd	Fwd	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t_{321}	0.384	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_{322}	0.385	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_{323}	0.387	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}	0.388	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.

TABLE 2.

The predata-mining *DS* of distance protective relay subjected to zone 1 A-G fault.

Here, *A* is $A = C \cup D$ which is a nonempty finite union set of condition and decision attributes (condition attributes $c_i \subset C$ suggest the multifunctional protective elements and analog measurands while decision attribute $d_i \subset D$ suggests the relay's trip output).

This big data is a hindrance in a laborious manual extraction of relay operation characteristics for the Expert System development. Thus, the aforementioned novel integrated data mining strategy is necessary to address this issue.

The resulting prepared decision table (after data selection, preprocessing, and transformation) of the distance protective relay's decision system is shown in **Table 2**. It is also called postdata-preparation*DS* or predata-mining *DS*. "." denotes data patterns that are similar to events immediately before and after them. Thus, they are not presented in order to reduce the table dimension. It is noticeable that the number of attributes has been substantially reduced by the data preparation strategy to merely 46 from the original 108 in the large raw event report.

The important analysis steps in the framework of Rough Set based data mining for deriving the distance relay decision algorithm from its event database is illustrated in **Figure 4** and discussed herewith.

The *computation of reducts* which is a process of reducing the number attributes while still maintaining the original data syntax is performed to start with. Within this the following substeps are executed:

a. Computation of the *D*-discernibility matrix of *C* (denoted as $\mathcal{M}_C(D)$). An element of $\mathcal{M}_C(D)$ is defined as the set of all condition attributes which discern events *ti* and *tj* and do not belong to the same equivalence class of the relation *U*|*IND*(*D*).

b. Subsequent derivation of the discernibility function *f*C(*D*) in Conjunctive Normal Form (CNF) (also called POS form in Boolean algebra) from *M*C(*D*). The CNF is reduced to final form after absorption law and omission of duplicates of disjunctive terms (sums) are applied minus the multiplication among each of the disjunctive terms of the final CNF.

c. In empirical database such as in this relay event data analysis, the calculation toward arriving at the final Disjunctive Normal Form (DNF) in order to find the eventual reducts is extremely computationally intensive. (DNF is obtained if the multiplication among each of the disjunctive terms of the final CNF is performed). In this case, the generation of reducts is considered as an NP-hard problem [19]. Thus, Genetic Algorithm is adopted to compute approximations of reducts by finding the minimally approximate hitting sets (analogous to reducts) from the sets corresponding to the discernibility function [20, 21].

Next *prediction rules* (denoted as $C \overset{pred}{\Longrightarrow} D$) are generated in which the above discovered reducts serve as the templates for the prediction rules to be created from. This is principally done by superimposing each reduct in the reduct set over the original decision table *DS* and then reading off the domain values of the condition and decision attributes. The resulting logical patterns, denoted as $C \Longrightarrow D$), that relate descriptions of condition to decision classes shall have the representation shown in Eq. (**1**):

$$C \overset{pred}{\Longrightarrow} D: IF\ C_i = v_{C_i}\ AND \ldots AND\ C_k = v_{C_k}\ THEN\ Trip = v_{Trip} \quad (1)$$

C⇒predD:IFci=vciAND…ANDck=vckTHENTrip=vTrip

Options

These prediction rules that are an exact representation of the characteristics of the relay decision system (table) *DS* can be described as the relay decision algorithm and can be designated as *ALG*(*DS*), i.e.,

$$ALG(DS) = \cup_{t \in \cup} (C \overset{pred}{\Longrightarrow} D)_t \quad (2)$$

ALG(DS)=∪t∈∪(C⇒predD)t

Options

where $(C \overset{pred}{\Longrightarrow} D)_t$(C⇒predD)t is the set of minimal prediction rules $C \overset{pred}{\Longrightarrow} D$C⇒predD for an event *t* ∈ ∪, i.e.,

$(C \overset{pred}{\Longrightarrow} D)_t$:IF$C_i$=$v_{ci}$(t)AND…AND$C_k$=$v_{ck}$(t)THENTrip=$v_{Trip}$(t)(C⇒predD)t:IFci=vci(t)AND…ANDck=vck(t)THEN Trip=vTrip(t) (3)

Options

This *ALG*(*DS*) can be evaluated for its accuracy as follows:

a. The entire original relay data set *DS* is partitioned into training and test sets using k-fold cross validation technique.

b. Estimating classification performance of the relay decision algorithm by rule firing-voting strategies.

The discovered *ALG*(*DS*) has been evaluated and verified by Othman et al. [15–17] to be able to be used to predict and discriminate future relay events having unknown trip state in unsupervised learning. This evaluation is necessary prior to allowing the eventual deduction of the relay association rule to take place.

Finally, postpruning (or filtering) is performed on the generated prediction rules $(C \overset{pred}{\Longrightarrow} D)$(C⇒predD) so as to discover relay *association rules* (denoted as $C \overset{pred}{\Longrightarrow} D$C⇒predD). These pertinent association rules essentially characterize the tripping decision logic of protective relay upon fault detection. This has been referred at the outset as the hypothesization of protective relay operation. This final version of knowledge representation shall be the main constituent for the Expert System knowledge base.

Because there are too large prediction rules to be filtered from, it is difficult to manually determine which rules are more useful, interesting, or important. Therefore, a measure of rule quality called *G2 Likelihood Ratio Statistic* as well as a measure of rule interestingness are used to select the most appropriate relay association rules and filter away the unwanted ones.

As mentioned above, these finally discovered relay association rules essentially describe the logical pattern of the correlating descriptions of conditions (i.e., *C*, the attribute set for various multifunctional protection elements) and the decision class (i.e., *D*, the attribute for trip assertion status). Thus, the symbol *CD* is used to illustrate *C*-*D* association and "*CD-association rule*" has been labeled as such to recognize it.

The final *CD*-association rule for one such fault condition as zone 1 A–G fault is shown in Eq. (4). Different fault condition would provide correspondingly different association rules to describe the relay's behavior.

IFZag(123)ANDCB52_A(closed)ANDpg_PkUp(123)ANDFltType(AGflt)ANDpp50_Z3(A)ANDpp50_Z4(A)ANDp50_Z1(A)AND p50_Z3(A)ANDr50(1234)ANDQ32(Fwd)ANDZload(0)ANDQ50(1234)ANDDist_ag(123)ANDpg_Trp(1)THENTrip(AIFZag(123)ANDCB52_A(closed)ANDpg_PkUp(123)ANDFltType(AGflt)ANDpp50_Z3(A)ANDpp50_Z4(A)ANDp50_Z1(A)AND p50_Z3(A)ANDr50(1234)ANDQ32(Fwd)ANDZload(0)ANDQ50(1234)ANDDist_ag(123)ANDpg_Trp(1)THENTrip(A (4)

Options

It is important to note that Eq. (4) defines the necessary triggering of the required relay multifunctional protective elements (antecedent) in order to recognize the zone 1 phase-A-to-ground fault and consequently assert the trip signal (consequent) to open pole A of the circuit breaker concerned. This is what the protection engineers would like to know in understanding the domain of the distance relay in responding to the fault.

Thus, it is necessary to verify how true it is that this rule can be used to interpret the distance relay behavior subjected to zone 1 A–G fault as represented by the predata-mining *DS* in Table 2. Out of all the relay events in the entire length of the relay event report, relay events *t90* and *t91* identified as the*fault detection* and *trip signal assertion* instances, respectively, will be our emphasis for cross reference to verify the exactness of the above-mentioned rationalized *CD*-association rule. In Table 2, the rule is seen to be an exact interpretation of the relay events *t90* and *t91*. Thus, the discovered rationalized *CD*-association rule is verified.

The eventually discovered $(C{=}{\Rightarrow}{=}_{\mathrm{assoc}}D)(C{\Rightarrow}\mathrm{assoc}D)$, and thus the desired hypothesis, has been proven to be an exact manifestation of the relay operation characteristics hidden in the event report [15–17]. The intelligent data mining framework provides the potential facility to conveniently discover exhaustively available knowledge of relay behavior from big event data subjected to exhaustively possible fault contingencies. Ultimately, a complete rule base for inference execution of an Expert System for relay operation analysis can be developed. This is the motivation of developing an Expert System called Protective Relay Analysis System (PRAY) that provides a platform for gathering previously discovered rules for its knowledge base construction.

3. Developing protective relay analysis system (PRAY) expert system

The concept of protective relay performance analysis is related to the convention that in any analysis known or correct events must first be hypothesized (expected operations are assumed), then an analysis is performed to confirm (validate) or refute the hypothesis by running matching exercise between expected and actual operations of the device under test [22]. If it is determined that the protective relay operation was incorrect, the diagnosis for cause must be performed [8]. This fundamental concept shall form the very basis of developing PRAY for distance protection.

PRAY is developed as an application tool under LabVIEW framework from National Instruments [23]. The main components of PRAY are as shown in **Figure 5** and described as follows:

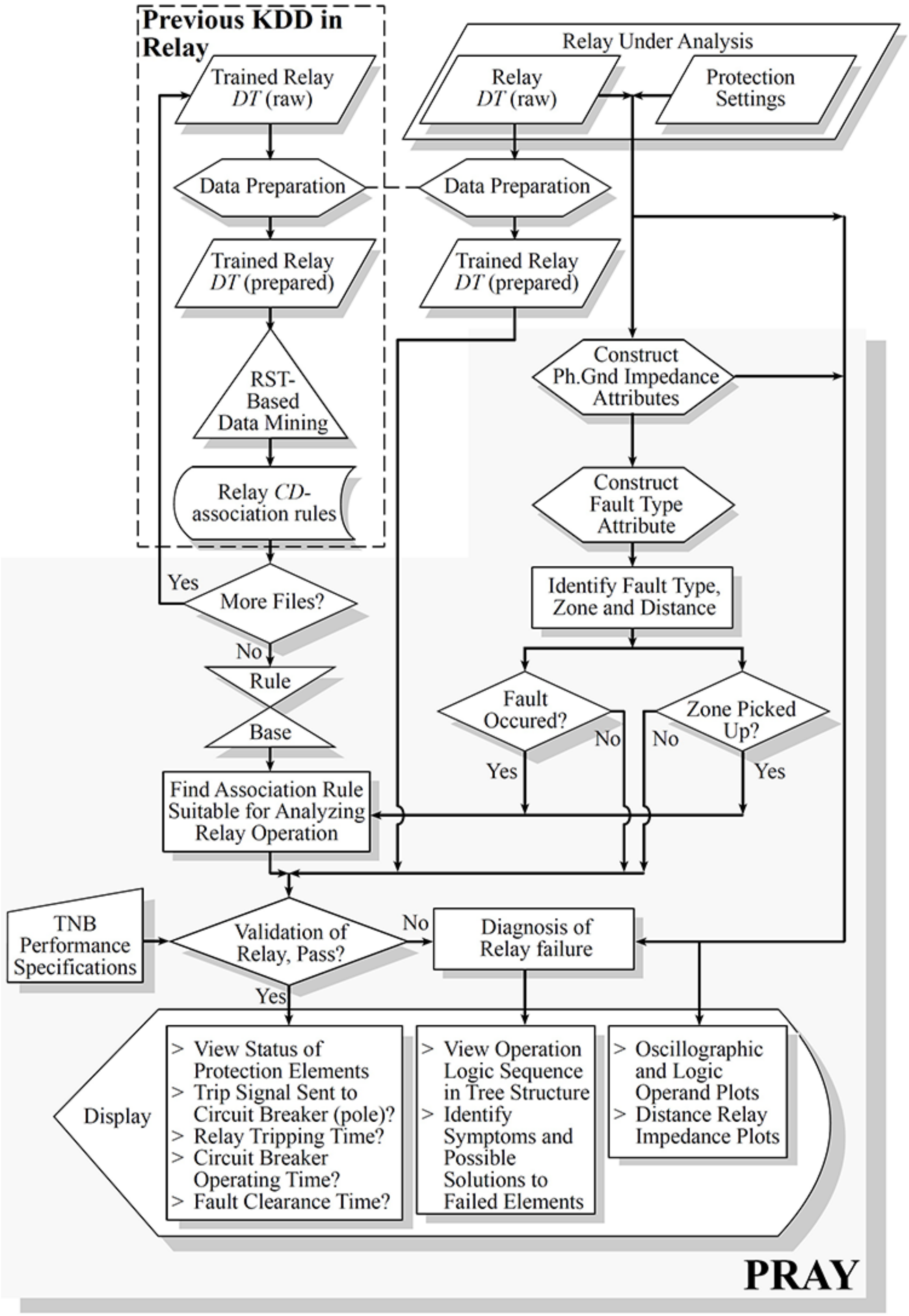

Previous KDD in Relay
Trained Relay DT (raw)
Data Preparation
Trained Relay DT (prepared)
RST-Based Data Mining
Relay CD-association rules
Relay Under Analysis
Relay DT (raw)
Protection Settings
Data Preparation
Trained Relay DT (prepared)
Construct Ph.Gnd Impedance Attributes
Construct Fault Type Attribute
Identify Fault Type, Zone and Distance
Fault Occured?
Zone Picked Up?
Yes
More Files?
No
Rule
Base
Find Association Rule Suitable for Analyzing Relay Operation
No
Yes
No
Yes
TNB Performance Specifications
Validation of Relay, Pass?
No
Diagnosis of Relay failure
Yes
Display
> View Status of Protection Elements
> Trip Signal Sent to Circuit Breaker (pole)?
> Relay Tripping Time?
> Circuit Breaker Operating Time?
> Fault Clearance Time?
> View Operation Logic Sequence in Tree Structure
> Identify Symptoms and Possible Solutions to Failed Elements
> Oscillographic and Logic Operand Plots
> Distance Relay Impedance Plots
PRAY

FIGURE 5.

Architecture of Protective Relay Analysis System (PRAY).

i. Construction of a rule base for PRAY’s inference engine by collating as an array all relay *CD*-association rules discovered from the KDD processes performed on trained relay. All attributes of each rule in the rule base shall be time tagged and arranged in a chronological order so that validation and diagnosis of the analyzed relay’s operations can be presented in an apparent operations logical sequence.

ii. Construction of phase and ground distance impedance channels (attributes) and fault-type channel. Using these channels, further identification processes of fault type, faulted zone, and distance to fault are executed and later used in singling out the most suitable relay *CD*-association rule from the rule base.

iii. Inferring, from the rule base according to both impending fault type and zone of pick-up, an expected relay *CD*-association rule to be best chosen as a hypothesis for the prediction of operations logic of the relay under analysis.

iv. Validation of occurrence of protective element pick-ups and their correctness of operations against hypothesis of the selected relay *CD*-association rule.

v. Symptom of relay element misoperation and its diagnosis as well as possible solution suggestion.

vi. Graphical plots of ground and phase impedance locus against respective ground and phase distance quadrilateral characteristics. The distance characteristics are constructed based on parameter settings taken from the relay under analysis. Instantaneous filtered voltages and currents and logic operands are also plotted.

3.1. PRAY INPUTS

The different inputs needed by PRAY for its analysis functions are as follows:

i. Relay *CD*-association rules: These rules saved as a plain text format in the KDD process are collated via graphical user interface (GUI) dialog input. The user is prompted for sufficient number of rules to be imported. The collated rules are converted into an array to form a rule base for the Expert System inference engine. Each rule input is an outcome of KDD after the Rough-Set-and-Genetic-Algorithm-based data mining and Rule Quality Measure (*G2* Likelihood Ratio Statistic) in ROSETTA [24]. In its untreated form, each rule input consists of a number of sub-*CD*-association rules. These subrules are rationalized into a single $C \Rightarrow D$ form by taking conjunction of them and using the concept of Boolean function manipulation by applying law of absorption.

ii. Analyzed relay event reports in the form of raw and prepared decision systems, (relay *DS*s): The raw relay *DS* is a converted data from relay resident IEEE COMTRADE format to DIAdem native format (.tdm), which is needed for processing in LabVIEW [25]. The prepared relay *DS* is a resultant file after the same data preparation process as that in the KDD for trained relay. This prepared relay *DS* in DIAdem format (.tdm) is of the same data structure as that used in the KDD; the latter is ready for the Rough Set data mining albeit not executed on for the expert system analysis. Having the same data structure is important so that the prepared *DS* of the relay under analysis can be correctly cross validated with a *CD*-association rule chosen from the PRAY rule base.

iii. Protection parameter settings: Imbedded as a separate “channel group” from the raw relay *DS’*s channel group in the same tdm file. The relay settings are originally recorded by the relay under analysis as a number of COMTRADE files. Since they are in the same file as the raw relay *DS*, they are also converted by DIAdem into tdm format.

iv. Performance specifications: The user has the option to key in values for parameters. For simplicity of analysis, TNB specifications for relay tripping time according to various zones of protection have been included as default values without requiring user’s inputs. (TNB is a short form for Tenaga Nasional Berhad, a Malaysian major utility organization.)

3.2. PRAY REASONING STRATEGY FOR VALIDATION AND DIAGNOSIS

The reasoning for validation and diagnosis of relay operations analysis starts with identification of fault type, faulted zone, and distance to fault by PRAY itself. The information from the fault type and picked-up faulted zone is then used to determine the index in the rule base array to determine the subarray containing the appropriate relay *CD*-association rule to be used in analyzing the relay under analysis. This chosen rule shall act as the hypothesis of anticipated operations of individual protective elements in the relay under analysis when a particular fault has occurred. All the antecedents and consequent in the rule have been initially arranged in sequential order during the rule base construction according to the time instances that have been tagged alongside them. Time tagging is important so that validation and diagnosis of relay operations can be executed according to the logical sequence stipulated by the hypothesis. This logical sequence is in fact indicative of relay operations logic. The following is a fictitious example of relay operation hypothesis based on a chosen relay *CD*-association rule:

- 0.000 *CB52_B*(closed) *Q32*(Fwd)
- 0.096 *p50_Z1*(B)
- 0.097 *FltType*(BGflt)
- 0.100 *Q50*(1234) *r50*(1234)
- 0.104 *Zload*(0)
- 0.107 *Dist_bg*(123) *Zbg*(123) *pg_PkUp*(123) *pg_Trp*(1)
- 0.108 *Trip*(B)

The consequent *Trip*(B) is associated with antecedents occurring beforehand. Any protective elements (antecedents) on the same row having the same time tagging indicate that they pick up (or stay in certain states) in concurrence. Expectedly, the last row having the highest tagged time must be the consequent (decision attribute) *Trip*(B).

The validation strategy of the operations of the analyzed relay starts by iterating through all antecedents in the hypothesis and comparing each one with that of the corresponding attribute of the prepared *DS* of the relay under analysis. Matched values result in messages describing the correctness of operations of the respective protective elements. On the other hand, any differences in the cross matches (either due to wrong pick-up values or nonassertion of the respective protective elements) will produce messages describing the relay's failed elements. The result of the validation is presented starting from the consequent (decision attribute, "*Trip*") at the top followed by antecedents arranged in descending sequence according to the order of the time tags in the hypothesis.

Diagnosis is carried out on failed, inoperative or misoperative protective elements. To view the cause–effect of events, a hierarchical tree is constructed based on the hypothesis where nodes are all hierarchically time sequenced, increasing in time from downstream nodes toward root node. The root node (top most) is the consequent of all the downstream antecedent nodes. Antecedents at the same nodes (i.e., having the same indentation) are concurrent in time instance. For the above-mentioned hypothesis, the diagnosis shall follow the following hierarchy:

Trip(B)

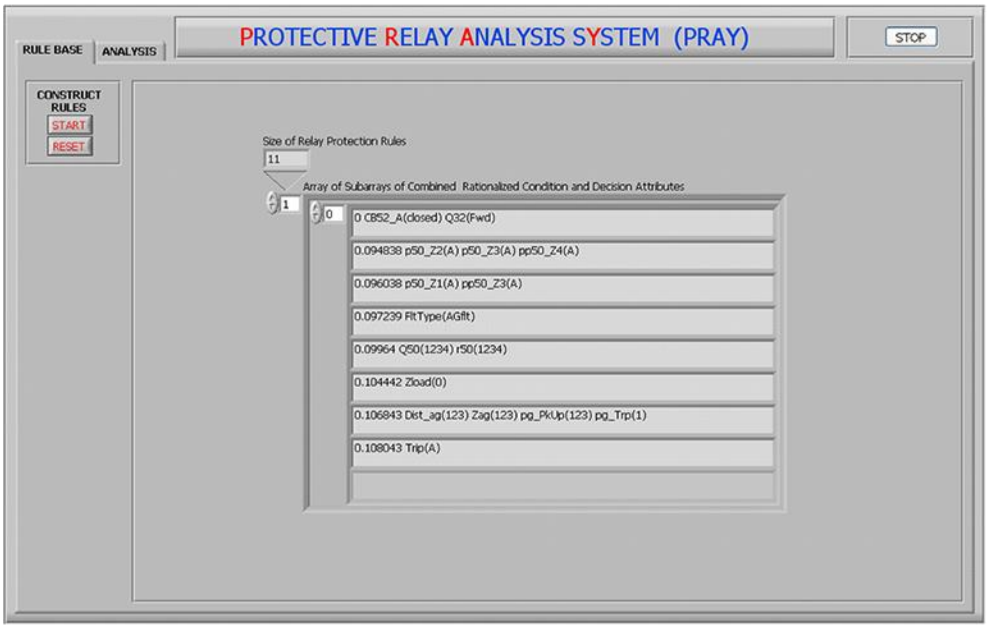
- - *Dist_bg*(123)
- - *Zbg*(123)
- - *pg_PkUp*(123)
- - *pg_Trp*(1)
-     \- *Zload*(0)
-         \- *Q50*(1234)

- - *r50*(1234)
- - *FltType*(BGflt)
- - *p50_Z1*(B)
- - *CB52_B*(closed)
- - *Q32*(Fwd)

4. PRAY analysis system results

In the rule base construction of PRAY, each of the imported *CD*-association rules, prior to being rationalized using the concept of Boolean function manipulation by applying the law of absorption, would be formatted by ROSETTA into a text file. When imported into PRAY, the file will be cleared of all unnecessary data such as comments and rule interestingness numerical measures leaving only the required relay *CD*-association rules for subsequent rationalization.

Figure 6 illustrates the GUI for the constructed rule base. Size of rule base and the selected subarray (0-indexed) of collated rule base array are shown. The size of the rule base reflects the number of training of various fault contingencies the trained relay has been subjected to.



Size of Relay Protection Rules
11

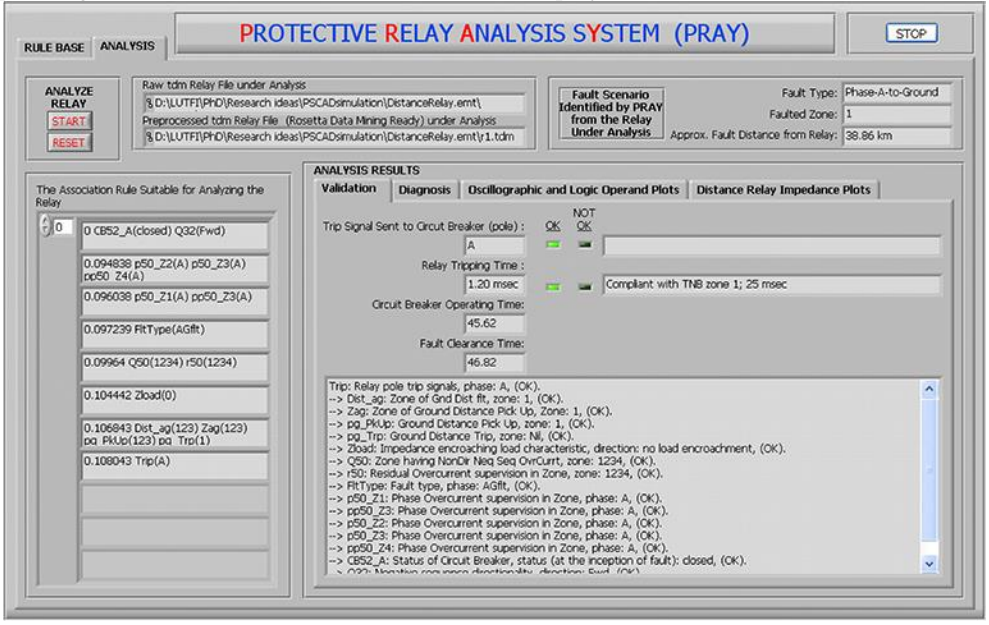
Array of Subarrays of Combined Rationalized Condition and Decision Attributes
1
0

```
0 CB52_A(closed) Q32(Fwd)
0.094838 p50_Z2(A) p50_Z3(A) pp50_Z4(A)
0.096038 p50_Z1(A) pp50_Z3(A)
0.097239 FltType(AGflt)
0.09964 Q50(1234) r50(1234)
0.104442 Zload(0)
0.106843 Dist_ag(123) Zag(123) pg_PkUp(123) pg_Trp(1)
0.108043 Trip(A)
```

FIGURE 6.

GUI for constructed rule base.

Figure 7 illustrates the GUI for analysis of a distance protective relay operation that has been subjected to a zone-1-AG fault. Using data in the relay’s raw tdm file, PRAY discovered that an AG fault has indeed occurred in zone 1 of the relay under analysis at approximately 39 km from its location in the substation. From this information, an appropriate relay *CD*-association rule has been chosen and displayed in the GUI. This rule shall be used to analyze whether any appropriate measures have been taken by the relay under analysis to clear the fault. In validating the individual operations of protective elements, the Validation field displays the correctness of actions taken by the relay after cross matching anticipated operations of individual protective elements hypothesized by the rule with the corresponding attributes obtained from the preprocessed tdm relay file under analysis. The consequent “Trip” is validated to have correctly sent a pole A trip signal to the circuit breaker. This is followed by correct antecedent statuses arranged in descending sequence according to the hypothesis. The relay tripping time of 1.2 ms is compliant with the TNB requirement of 25 ms for zone 1 operation. The circuit breaker operating time and fault clearance time are also displayed in the GUI.



RULE BASE
ANALYSIS
PROTECTIVE RELAY ANALYSIS SYSTEM (PRAY)
STOP
ANALYZE RELAY
START
RESET
ANALYSIS RESULTS
Validation
Diagnosis
Oscillographic and Logic Operand Plots
Distance Relay Impedance Plots
OK
NOT OK

Raw tdm Relay File under Analysis: D:\LUTFI\PhD\Research ideas\PSCADsimulation\DistanceRelay.emt\

Preprocessed tdm Relay File (Rosetta Data Mining Ready) under Analysis: D:\LUTFI\PhD\Research ideas\PSCADsimulation\DistanceRelay.emt\r1.tdm

Fault Scenario Identified by PRAY from the Relay Under Analysis

Fault Type: Phase-A-to-Ground

Faulted Zone: 1

Approx. Fault Distance from Relay: 38.86 km

The Association Rule Suitable for Analyzing the Relay: 0

0 CB52_A(closed) Q32(Fwd)

0.094838 p50_Z2(A) p50_Z3(A) pp50_Z4(A)

0.096038 p50_Z1(A) pp50_Z3(A)

0.097239 FltType(AGflt)

0.09964 Q50(1234) r50(1234)

0.104442 Zload(0)

0.106843 Dist_ag(123) Zag(123) pg_PkUp(123) pg_Trp(1)

0.108043 Trip(A)

Trip Signal Sent to Circut Breaker (pole) : A

Relay Tripping Time : 1.20 msec — Compliant with TNB zone 1; 25 msec

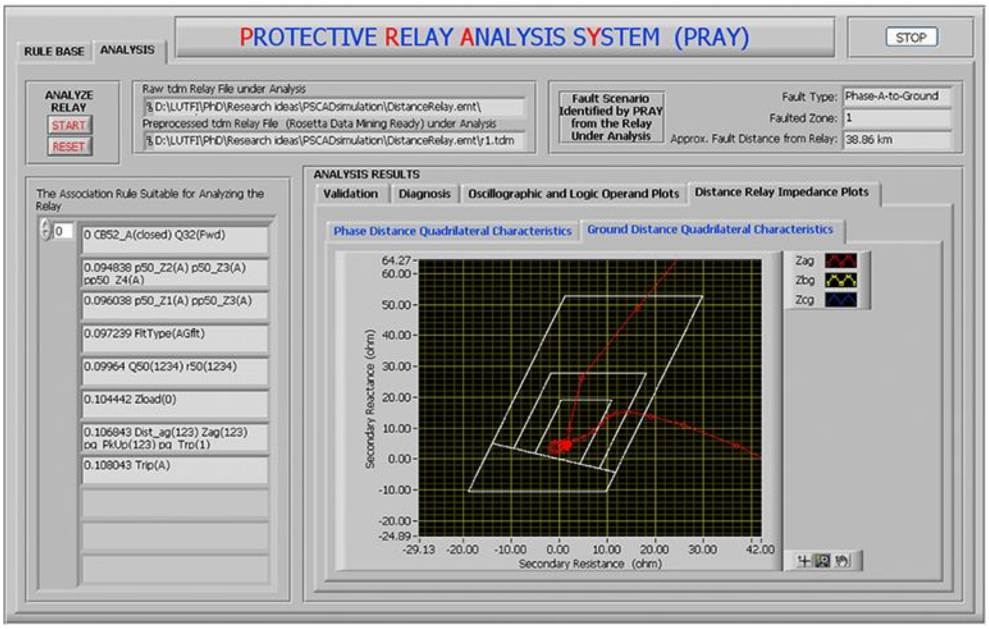
Circuit Breaker Operating Time: 45.62

Fault Clearance Time: 46.82

```
Trip: Relay pole trip signals, phase: A, (OK).
--> Dist_ag: Zone of Gnd Dist flt, zone: 1, (OK).
--> Zag: Zone of Ground Distance Pick Up, Zone: 1, (OK).
--> pg_PkUp: Ground Distance Pick Up, zone: 1, (OK).
--> pg_Trp: Ground Distance Trip, zone: Nil, (OK).
--> Zload: Impedance encroaching load characteristic, direction: no load encroachment, (OK).
--> Q50: Zone having NonDir Neq Seq OvrCurrt, zone: 1234, (OK).
--> r50: Residual Overcurrent supervision in Zone, zone: 1234, (OK).
--> FltType: Fault type, phase: AGflt, (OK).
--> p50_Z1: Phase Overcurrent supervision in Zone, phase: A, (OK).
--> pp50_Z3: Phase Overcurrent supervision in Zone, phase: A, (OK).
--> p50_Z2: Phase Overcurrent supervision in Zone, phase: A, (OK).
--> p50_Z3: Phase Overcurrent supervision in Zone, phase: A, (OK).
--> pp50_Z4: Phase Overcurrent supervision in Zone, phase: A, (OK).
--> CB52_A: Status of Circuit Breaker, status (at the inception of fault): closed, (OK).
--> Q32: Negative sequence directionality, direction: Fwd, (OK).
```

FIGURE 7.

GUI for analysis of distance protective relay operations.



RULE BASE
ANALYSIS
PROTECTIVE RELAY ANALYSIS SYSTEM (PRAY)
STOP
ANALYZE RELAY
START
RESET
ANALYSIS RESULTS
Validation
Diagnosis
Oscillographic and Logic Operand Plots
Distance Relay Impedance Plots
Phase Distance Quadrilateral Characteristics
Ground Distance Quadrilateral Characteristics
64.27
60.00
50.00
40.00
30.00
20.00
10.00
0.00
-10.00
-20.00
-24.89
Secondary Reactance (ohm)
-29.13
-20.00
-10.00
0.00
10.00
20.00
30.00
42.00
Secondary Resistance (ohm)
Zag
Zbg
Zcg

Raw tdm Relay File under Analysis: D:\LUTFI\PhD\Research ideas\PSCADsimulation\DistanceRelay.emt\

Preprocessed tdm Relay File (Rosetta Data Mining Ready) under Analysis: D:\LUTFI\PhD\Research ideas\PSCADsimulation\DistanceRelay.emt\r1.tdm

Fault Scenario Identified by PRAY from the Relay Under Analysis

Fault Type: Phase-A-to-Ground

Faulted Zone: 1

Approx. Fault Distance from Relay: 38.86 km

The Association Rule Suitable for Analyzing the Relay: 0

0 CB52_A(closed) Q32(Fwd)

0.094838 p50_Z2(A) p50_Z3(A) pp50_Z4(A)

0.096038 p50_Z1(A) pp50_Z3(A)

0.097239 FltType(AGflt)

0.09964 Q50(1234) r50(1234)

0.104442 Zload(0)

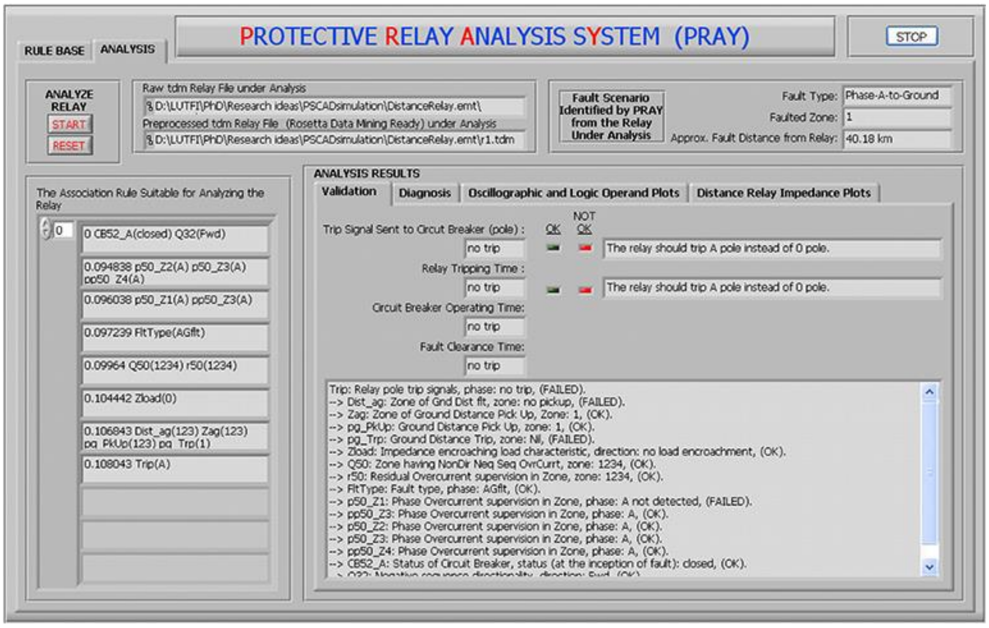
0.106843 Dist_ag(123) Zag(123) pg_PkUp(123) pg_Trp(1)

0.108043 Trip(A)

FIGURE 8.

GUI for ground distance quadrilateral characteristics plots.

Figure 8 shows the graphical plots of ground impedance locus against respective ground distance quadrilateral characteristics. Since the fault is AG occurring in zone 1, it is noted that only trajectory of *Zag* traverses through into zone 1 of the ground quadrilateral characteristics and all phase impedances stay as outliers of the phase quadrilateral characteristics as expected.


PROTECTIVE RELAY ANALYSIS SYSTEM (PRAY)
RULE BASE
ANALYSIS
STOP
ANALYZE RELAY
START
RESET
ANALYSIS RESULTS
Validation
Diagnosis
Oscillographic and Logic Operand Plots
Distance Relay Impedance Plots

Raw tdm Relay File under Analysis: D:\LUTFI\PhD\Research ideas\PSCADsimulation\DistanceRelay.emt\

Preprocessed tdm Relay File (Rosetta Data Mining Ready) under Analysis: D:\LUTFI\PhD\Research ideas\PSCADsimulation\DistanceRelay.emt\r1.tdm

Fault Scenario Identified by PRAY from the Relay Under Analysis

Fault Type: Phase-A-to-Ground

Faulted Zone: 1

Approx. Fault Distance from Relay: 40.18 km

The Association Rule Suitable for Analyzing the Relay: 0

- 0 CB52_A(closed) Q32(Fwd)
- 0.094838 p50_Z2(A) p50_Z3(A) pp50_Z4(A)
- 0.096038 p50_Z1(A) pp50_Z3(A)
- 0.097239 FltType(AGflt)
- 0.09964 Q50(1234) r50(1234)
- 0.104442 Zload(0)
- 0.106843 Dist_ag(123) Zag(123) pg_PkUp(123) pg_Trp(1)
- 0.108043 Trip(A)

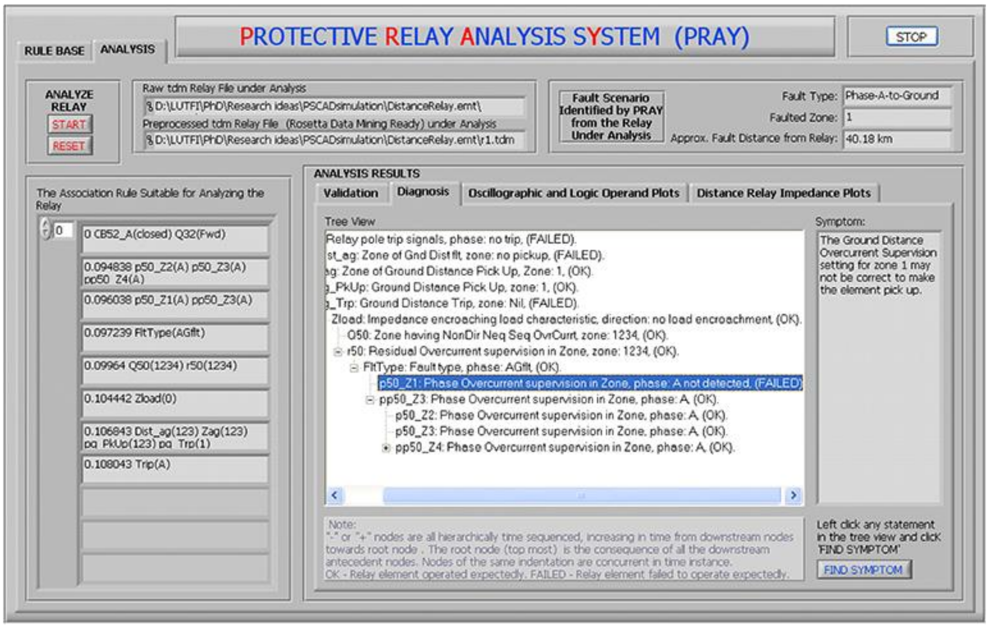
		OK	NOT OK	
Trip Signal Sent to Circut Breaker (pole) :	no trip		🟥	The relay should trip A pole instead of 0 pole.
Relay Tripping Time :	no trip		🟥	The relay should trip A pole instead of 0 pole.
Circuit Breaker Operating Time:	no trip			
Fault Clearance Time:	no trip			

```
Trip: Relay pole trip signals, phase: no trip, (FAILED).
--> Dist_ag: Zone of Gnd Dist flt, zone: no pickup, (FAILED).
--> Zag: Zone of Ground Distance Pick Up, Zone: 1, (OK).
--> pg_PkUp: Ground Distance Pick Up, zone: 1, (OK).
--> pg_Trp: Ground Distance Trip, zone: Nil, (FAILED).
--> Zload: Impedance encroaching load characteristic, direction: no load encroachment, (OK).
--> Q50: Zone having NonDir Neq Seq OvrCurrt, zone: 1234, (OK).
--> r50: Residual Overcurrent supervision in Zone, zone: 1234, (OK).
--> FltType: Fault type, phase: AGflt, (OK).
--> p50_Z1: Phase Overcurrent supervision in Zone, phase: A not detected, (FAILED).
--> pp50_Z3: Phase Overcurrent supervision in Zone, phase: A, (OK).
--> p50_Z2: Phase Overcurrent supervision in Zone, phase: A, (OK).
--> p50_Z3: Phase Overcurrent supervision in Zone, phase: A, (OK).
--> pp50_Z4: Phase Overcurrent supervision in Zone, phase: A, (OK).
--> CB52_A: Status of Circuit Breaker, status (at the inception of fault): closed, (OK).
--> Q32: Negative sequence directionality, direction: Fwd, (OK).
```

FIGURE 9.

Validation of misoperative relay.

Figure 9 illustrates a screenshot of PRAY's validation for a distance relay that had failed to operate (maloperated) when the transmission line it was protecting was subjected to a zone-1-AG fault. PRAY discovered that an AG fault had occurred in one of the relays under analysis at approximately 40 km forward its location in the substation. (This is actually the same fault occurred in the above analysis of the same relay operating successfully.) From this information, an appropriate relay *CD*-association rule had been chosen as the hypothesis (similar to the above) and used to validate that appropriate measures had not been taken to clear the fault. The consequent "Trip" was validated to have not sent a pole-A trip signal to the circuit breaker. The descending sequence of antecedents indicated that although there were correct operations of negative sequence overcurrent (*Q50*) and residual overcurrent supervision (*r50*) elements, signifying the impending A–G imbalanced fault, the zone-1 overcurrent supervision element (*p50_Z1*) had failed to do likewise. This was believed to have attributed to the relay's failure to trip. Looking at the operation logic of different protective elements at different levels of sequence in the Diagnosis field's hierarchical tree, it is apparent that the failure by the overcurrent element *p50_Z1*is diagnosed to be the possible cause of the relay maloperation. Finding the symptom related to the malfunctional *p50_Z1* element as shown in **Figure 10** reveals that an incorrect threshold setting could have caused its failure.


RULE BASE
ANALYSIS
PROTECTIVE RELAY ANALYSIS SYSTEM (PRAY)
STOP
ANALYZE RELAY
START
RESET
ANALYSIS RESULTS
Validation
Diagnosis
Oscillographic and Logic Operand Plots
Distance Relay Impedance Plots
FIND SYMPTOM

Raw tdm Relay File under Analysis: %D:\LUTFI\PhD\Research ideas\PSCADsimulation\DistanceRelay.emt\

Preprocessed tdm Relay File (Rosetta Data Mining Ready) under Analysis: %D:\LUTFI\PhD\Research ideas\PSCADsimulation\DistanceRelay.emt\r1.tdm

Fault Scenario Identified by PRAY from the Relay Under Analysis

Fault Type: Phase-A-to-Ground

Faulted Zone: 1

Approx. Fault Distance from Relay: 40.18 km

The Association Rule Suitable for Analyzing the Relay

0

0 CB52_A(closed) Q32(Fwd)

0.094838 p50_Z2(A) p50_Z3(A) pp50_Z4(A)

0.096038 p50_Z1(A) pp50_Z3(A)

0.097239 FltType(AGflt)

0.09964 Q50(1234) r50(1234)

0.104442 Zload(0)

0.106843 Dist_ag(123) Zag(123) pg_PkUp(123) pg_Trp(1)

0.108043 Trip(A)

Tree View

```
Relay pole trip signals, phase: no trip, (FAILED).
st_ag: Zone of Gnd Dist flt, zone: no pickup, (FAILED).
ag: Zone of Ground Distance Pick Up, Zone: 1, (OK).
_PkUp: Ground Distance Pick Up, zone: 1, (OK).
_Trp: Ground Distance Trip, zone: Nil, (FAILED).
 Zload: Impedance encroaching load characteristic, direction: no load encroachment, (OK).
   Q50: Zone having NonDir Neq Seq OvrCurrt, zone: 1234, (OK).
 ⊟ r50: Residual Overcurrent supervision in Zone, zone: 1234, (OK).
    ⊟ FltType: Fault type, phase: AGflt, (OK).
         p50_Z1: Phase Overcurrent supervision in Zone, phase: A not detected, (FAILED)
      ⊟ pp50_Z3: Phase Overcurrent supervision in Zone, phase: A, (OK).
           p50_Z2: Phase Overcurrent supervision in Zone, phase: A, (OK).
           p50_Z3: Phase Overcurrent supervision in Zone, phase: A, (OK).
         ⊞ pp50_Z4: Phase Overcurrent supervision in Zone, phase: A, (OK).
```

Symptom: The Ground Distance Overcurrent Supervision setting for zone 1 may not be correct to make the element pick up.

Note: "-" or "+" nodes are all hierarchically time sequenced, increasing in time from downstream nodes towards root node . The root node (top most) is the consequence of all the downstream antecedent nodes. Nodes of the same indentation are concurrent in time instance. OK - Relay element operated expectedly. FAILED - Relay element failed to operate expectedly.

Left click any statement in the tree view and click 'FIND SYMPTOM'

FIGURE 10.

Diagnosis of misoperative relay.

5. Summary

The developed Protective Relay Analysis (PRAY) Expert System has demonstrated how the problems related to the maintenance of rule base of an Expert System can be addressed. By collating all the necessary relay *CD*-association rules discovered previously from the earlier KDD processes involving integrated-Rough-Set-and-Genetic-Algorithm data mining, Rule Quality Measure, and rule interestingness and importance judgments (as discussed in the authors' cited works), a maintainable knowledge base for inference strategy can be conveniently prepared. Although this study revolves around analyzing a modeled distance relay's big event data by hypothesis discovery, validation, and diagnosis, it is envisaged that using this approach a more rigorous analysis implementation of actual protective relay of different types can be embarked on.

6. Acknowledgements

This work was supported by the Universiti Putra Malaysia under the Geran Putra IPB scheme with the project no. GP-IPB/2013/9412101.

Nomenclature

C	rule condition attribute(s)
CB52_B	status of circuit breaker.
C ⇒ *D*	relay decision rule, general term for (C=⇒=$_{\text{assoc}}$D)(C⇒assocD) and (C=⇒$_{\text{pred}}$D)(C⇒predD)
(C=⇒=$_{\text{assoc}}$D)(C⇒assocD)	relay *CD*-association rule
(C=⇒$_{\text{pred}}$D)(C⇒predD)	relay *CD*-prediction rule

CD-association rule	a relay association rule associating between *C* and *D*
CD-decision alg.	a set of relay prediction rules that predict *D* from *C* (alg. is algorithm)
CD-prediction rule	rule that predicts *D* from *C*
CNF	conjunctive normal form (i.e., product of sum (POS) in Boolean algebra).
COMTRADE	common format for transient data exchange, an IEEE file format
D	rule decision attribute
Dist_bg	zone of Gnd Dist flt (ground distance fault)
DNF	disjunctive normal form (i.e., sum of product (SOP) in Boolean algebra)
DS/DT	decision system/decision table
$f_C(D)$	discernibility function
FltType	fault type
GA	genetic algorithm
G2	*G2* Likelihood ratio statistic, a rule quality measure
IS	information system
KDD	Knowledge discovery in database
$M_C(D)$	*D*-discernibility matrix of *C*
p50_Z1	phase overcurrent supervision in zone
pg_PkUp	ground distance pick-up
pg_Trp	ground distance trip
PRAY	Protective relay analysis system, an Expert System
Q32	negative sequence directionality
Q50	zone having NonDir Neq Seq OvrCurrt (nondirectional negative sequence overcurrent)

r50	residual overcurrent supervision in zone
$RED_D(C)$	*D*-reducts of *C*, sets of reduced number of indispensable attributes
RST	Rough set theory
$M\ (f_C(D))$	multiset
$M\ (f_C(D))_{\text{Min Hit Set}}$	minimal hitting set
SOP	sum of products
Trip	relay pole trip signals
$U\|IND(D)$	indiscernibility-relation/equivalence-class/elementary-sets about universe of relay events *U*with respect to *D*
Zbg	zone of ground distance pick-up.
Zload	impedance encroaching load characteristic