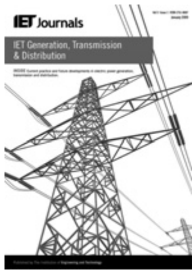
www.ietdl.org

Published in IET Generation, Transmission & Distribution
Received on 26th June 2013
Revised on 11th June 2014
Accepted on 14th June 2014
doi: 10.1049/iet-gtd.2014.0023



ISSN 1751-8687

Decision tree-induced fuzzy rule-based differential relaying for transmission line including unified power flow controller and wind-farms

Manas Kumar Jena, Subhransu Ranjan Samantaray, Lokanath Tripathy

School of Electrical Sciences, Indian Institute of Technology Bhubaneswar, India
E-mail: kumarmanas87@yahoo.com

Abstract: The paper presents a decision tree (DT) induced fuzzy rule-based intelligent differential relaying scheme for transmission lines including unified power flow controller (UPFC) and wind-farms. The conventional distance relaying scheme fails miserably in protecting transmission lines including flexible AC transmission systems controllers such as UPFC and further, the protection issues become more challenging when wind-farm is integrated. The proposed protection scheme extracts features from the instantaneous voltage and current signals from both ends of the transmission line using discrete Fourier transform based pre-processor and computes corresponding differential features. The differential features are used to build the fault classification tree using a data-mining algorithm known as DT. Further, the fuzzy membership functions are drawn using the DT thresholds and the corresponding fuzzy rule-base is developed for final relaying decision. The proposed technique has been extensively tested (in MATLAB environment) for single circuit and double circuit transmission lines including UPFC and wind-farms with wide variations in operating conditions. The test results indicate that the proposed DT-induced fuzzy rule-based relaying scheme is highly reliable and robust for protection of complex transmission line including UPFC and wind-farm.

1 Introduction

The relaying aspect of transmission lines including flexible AC transmission systems (FACTS) controllers [1], is one of the most challenging tasks for the power engineers in the modern era of integrated power systems. Among the different types of FACTS controllers available, unified power flow controller (UPFC) is considered to be one of the most effective devices for power flow control. It can provide simultaneous and independent control of important power system parameters; line active power flow, line reactive power flow, impedance and voltage. It offers necessary functional flexibility for the combined application of phase angle control with controlled series and shunt compensation and a versatile alternative to conventional reinforcement methods. However, because of the presence of FACTS controllers like UPFC in a fault loop, the voltage and current signals at the relay point are affected in steady state and transient state [2], which in turn affects the performance of the existing protection schemes. In present day power system operation, wind-farms are also increasingly integrated to the grids at different levels of voltage across the world. The share of such farms in the transmission system is rising day by day. The difficulty that arises in integrating such farms is primarily because of the uncontrollable wind speed. The transmission system that connects wind-farm is exposed to such a continuously changing environment. Thus, the protection task becomes more challenging [3, 4]. Wind-farm integration to the transmission line may also bring problems such as weak feed or weak source condition. The fault current in case of phase faults depends on the amount of generation at the instance of fault and the fault contribution characteristics of the machines. Some machines like doubly fed induction generator (DFIG) contributes only about 1.1 pu (110% of full generation), after a few cycles of the fault inception, resulting into a very weak source. Further, the fault current contribution of wind turbines with crowbar protected DFIG affects the performance of existing current differential and pilot protection schemes [5, 6]. When both UPFC and wind-farms are integrated together in the transmission lines, the system becomes more complicated and the performance of the conventional relaying scheme is greatly affected. In [7], performance of conventional distance relays in presence of UPFC compensated transmission lines have been investigated. It is clear from [7] that there is a need to develop a different relaying algorithm for transmission system compensated with UPFC. In [8], support vector machine (SVM) based fault classification scheme for compensated transmission line has been proposed. However, the computational time of SVM is higher compared to the proposed decision tree (DT) based data-mining algorithm. This puts constraint on the online realisation of SVM-based relays for relaying applications, where speed and accuracy are prime considerations. DT-based fault classification scheme for single circuit transmission lines is presented in [9]. The scheme does not include the effect of FACTS controllers upon the dependability of the algorithm. Thus, there is a need to develop a dedicated relaying strategy for transmission lines including UPFC and wind-farms together.

Data-mining has become a very successful research tool in recent times in power system applications such as power quality classification, islanding detection, fault classification etc. [10, 11]. Neural network (NN), SVM and decision tree (DT) are part of data-mining tool. The disadvantage associated with NN and SVM is that both of them are black-box solutions, whereas DT [12] is a more transparent solution. However, DT possesses sharp decision boundaries at the different nodes of the classification tree. The difficulty associated with the sharp partition of the continuous attributes is that a small change in attribute values of a case being classified may result in radical changes in classification. In certain circumstances this may result in misclassification, which can be avoided if the decision boundaries are more gradual. Thus, fuzzy theoretical techniques are used to fuzzify the crisp decision trees in order to soften the sharp decision boundaries at the decision nodes. Fuzzy-induced DT develops flexible boundaries for each decision threshold originally generated from the DT and avoids rectangular partitioning [13–17]. Application of fuzzified-DT to power systems was first proposed in [10] for security assessment in a large power network with trapezoidal fuzzy membership function. However, the proposed DT-induced fuzzy rule base includes triangular membership functions which has better uncertainty handling capacity and thus, able to build a more generalised and accurate classifier. Initially, the proposed technique pre-processes the instantaneous current and voltage signals retrieved at each end of the transmission line using the discrete Fourier transform (DFT) based pre-processor. Different features are computed at both ends of the line including UPFC and wind-farms [18] and further, the corresponding differential features are computed. Once the differential features are available, the corresponding fuzzy-induced DT is developed for final fault classification. The proposed algorithm is tested for all ten types of shunt faults with wide variations in fault as well as system parameters. The following sections deal with proposed relaying scheme, initial system studied, results analysis, discussion and conclusions.

2 Proposed relaying scheme

The proposed relaying scheme computes current and voltage phasors from the instantaneous current and voltage signals using DFT-based phasor extractor at both ends of the transmission line. The following operations take place inside the phasor extractor module: (i) analogue signal sampling and DC-offset removal; (ii) DFT extraction of fundamental phasor. The sampling frequency of the proposed scheme is 4800 Hz (96 samples per cycle). The sampled data is then buffered into a moving average filter. The moving average filter is used to reduce random noise (white noise), while still providing a sharp step response. A full cycle DFT is used to calculate the phasors.

Once the current and voltage phasors are available, differential features are extracted which are used to train the DT. Further, the decision nodes of developed DT are fuzzified, which ultimately provides the final relaying decision. Different features based on sequence components, reactive power and the rate of change of the same are derived at both ends of the line and corresponding differential features are extracted which are used to build the DT. The flow chart of the proposed scheme is shown in Fig. 1, which includes deriving differential features, building DT, transforming DT to fuzzy-DT and final relaying decision. The proposed scheme includes 21 possible differential features which could be mostly affected during the fault process [2, 8–11]. The differential features considered in the proposed study are as follows:

- $X_1 = \partial(V_{s1} - V_{r1})/dt$: (rate of change of positive sequence voltage difference)
- $X_2 = \partial(I_{s1} - I_{r1})/dt$: (rate of change of positive sequence current difference)
- $X_3 = \partial(V_{s2} - V_{r2})/dt$: (rate of change of negative-sequence voltage difference)
- $X_4 = \partial(I_{s2} - I_{r2})/dt$: (rate of change of negative-sequence current difference)
- $X_5 = \partial(V_{s0} - V_{r0})/dt$: (rate of change of zero-sequence voltage difference)
- $X_6 = \partial(I_{s0} - I_{r0})/dt$: (rate of change of zero-sequence current difference)
- $X_7 = \partial(V_{sa} - V_{ra})/dt$: (rate of change of phase a voltage difference)
- $X_8 = \partial(V_{sb} - V_{rb})/dt$: (rate of change of phase b voltage difference)
- $X_9 = \partial(V_{sc} - V_{rc})/dt$: (rate of change of phase c voltage difference)
- $X_{10} = (Q_{sa} - Q_{ra})$: (reactive power difference for phase-a)
- $X_{11} = (Q_{sb} - Q_{rb})$: (reactive power difference for phase-b)
- $X_{12} = (Q_{sc} - Q_{rc})$: (reactive power difference for phase-c)
- $X_{13} = (V_{s1} - V_{r1})$: (positive sequence voltage difference)
- $X_{14} = (V_{s2} - V_{r2})$: (negative-sequence voltage difference)
- $X_{15} = (V_{s0} - V_{r0})$: (zero-sequence voltage difference)
- $X_{16} = (I_{s1} - I_{r1})$: (positive sequence current difference)
- $X_{17} = (I_{s2} - I_{r2})$: (negative-sequence current difference)
- $X_{18} = (I_{s0} - I_{r0})$: (zero-sequence current difference)
- $X_{19} = \partial(I_{sa} - I_{ra})/dt$: (rate of change of phase-a current)
- $X_{20} = \partial(I_{sb} - I_{rb})/dt$: (rate of change of phase-b current)
- $X_{21} = \partial(I_{sc} - I_{rc})/dt$: (rate of change of phase-c current)

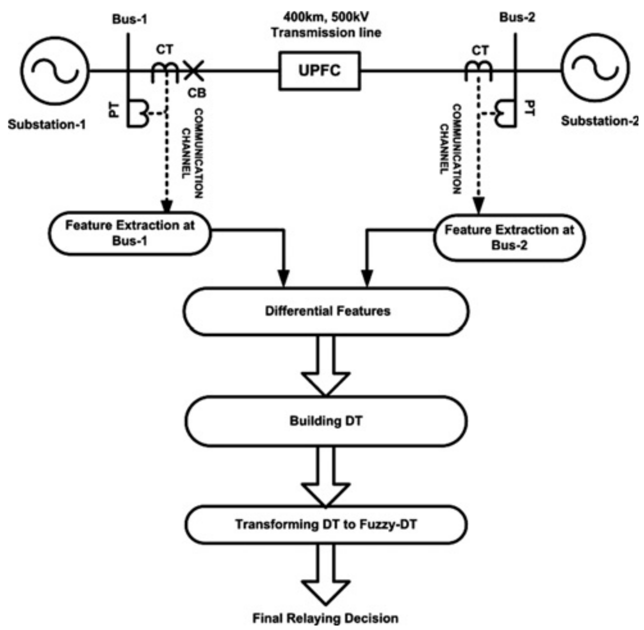


Fig. 1 *Schematic representation of proposed Relaying scheme*

Where subscripts 1, 2, 0 are used to indicate positive, negative and zero-sequence components, respectively. Subscripts 's' and 'r' are used to indicate sending (substation-1) and receiving end (substation-2) of the studied systems, respectively. The sampled instantaneous voltage and current signals are fed to the DFT pre-processor, which is developed using MATLAB (SIMULINK). The DFT-based feature extractor computes 21 features. The input for the DT is a vector consists of 21 data points against a particular target output (one type of fault). The target outputs (classes) are categorised as 0 (no fault/external fault), 1(*a–g*), 2(*b–g*), 3(*c–g*), 4(*a–b*), 5(*b–c*), 6(*c–a*), 7(*ab–g*) 8(*bc–g*), 9(*ca–g*), 10(*a–b–c*). The Rattle software package [19] is used in the proposed study to develop DT for classifying different types of faults for the system studied.

The proposed study considers wide variation in operating conditions and system parameters as follows.

- Variation in fault resistance (R_f) from 0 to 100 Ω
- Variation in fault location: 20, 30, 50,70, 80, 85 and 90% of the total line length
- Variation in fault inception angle(FIA): 0°,30°,60°,90°
- Variations in source impedance angle: 30% from normal value.
- Different types of fault: *a–g*, *b–g*, *c–g*, *a–b*, *b–c*, *c–a*, *ab–g*, *bc–g*, *ca–g*, *a–b–c*
- UPFC series injected voltage (V_{se}) varied for 0–15% of the normal voltage
- UPFC voltage phase angle(θ_{se}) varied from 0°–360°
- UPFC control mode (automatic power flow control mode and bypass mode)
- Variation in wind speed: 10, 15, 20 m/s
- Reverse power flow
- Remote in-feed
- Noisy environment (signal-to-noise ratio: SNR 20 dB)

The complete data set generated considering above variations are used to train and test the DT. Different combinations of training and testing ratios are considered to check the patterns for best classification accuracy. The proposed study considers two systems for initial testing: (i) single circuit transmission line with both UPFC and wind-farm; and (ii) double circuit transmission line with UPFC and wind-farm

The proposed intelligent relaying scheme requires a reliable long-distance communication channel to exchange synchronised voltage and current measurements measured at two end of the line. The newly built extra high voltage/ultra high voltage (EHV/UHV) transmission lines [20] is well equipped with dedicated fibre-optics channels, through which the three-phase voltages and currents can be transmitted from one end of the transmission line to the other.

3 Building DT for initial system studied

A 500 kV, 50 Hz power system is illustrated in Fig. 2 (test case I: single circuit transmission line with UPFC and wind-farm). This studied system has two substations (sending end and receiving end) and one UPFC located at the mid-point of the transmission line (distributed model). Wind-farm is connected at the receiving end of the studied system. Hence, the system consists of two sources, UPFC and its associated components and a 400 km transmission line. The transmission line parameters are as follow:

$R_1 = 0.01537$ Ω/km: positive sequence resistance
$R_0 = 0.04612$ Ω/km: zero-sequence resistance
$R_{0\,m} = 0.20052$ Ω/km: zero-sequence mutual resistance (for double circuit line)
$L_1 = 0.8858 \times 10^{-3}$ H/km: positive sequence inductance
$L_0 = 2.654 \times 10^{-3}$ H/km: zero-sequence inductance
$L_{0\,m} = 0.0020802$ H/km: zero-sequence mutual inductance (for double circuit line)
$C_1 = 13.06 \times 10^{-9}$ F/km: positive sequence capacitance
$C_0 = 4.355 \times 10^{-9}$ F/km: zero-sequence capacitance
$C_{0\,m} = -2.0444\text{e}^{-9}$ F/km: zero-sequence mutual capacitance (for double circuit line) and
$V_s = 500\ \text{kV}\angle\delta_1$ (substation-1)
$V_r = 500\ \text{kV}\angle\ \delta_2$ (substation-2)
$V_3 = 500\ \text{kV}\angle\ \delta_3$ (substation-3 for in-feed line)

A 100 MVA UPFC is installed in the middle of the transmission line of length 400 km (i.e. 200 km from relaying end). UPFC [1] consists of two 48-pulse voltage source inverters which are connected through two 2500 µF common DC capacitors. The first inverter, known as static compensator, connects into the transmission system through a 15 kV/500 kV Δ/ϒ shunt transformer and injects or consumes reactive power to the transmission system to regulate the voltage at the connecting point. Another inverter, known as static synchronous series compensator, connects into the system through a 15 kV/22 kV ϒ/ϒ series transformer to inject an almost sinusoidal voltage of variable magnitude and angle in series with the transmission line to regulate the power flow through the line. The details of the UPFC modelling considered in our proposed relaying scheme are derived from [2].

The wind-farm consists of 40 numbers of 1.5 MW wind turbines connected to the 500 kV system. Wind turbines use a DFIG consisting of a wound rotor induction generator and

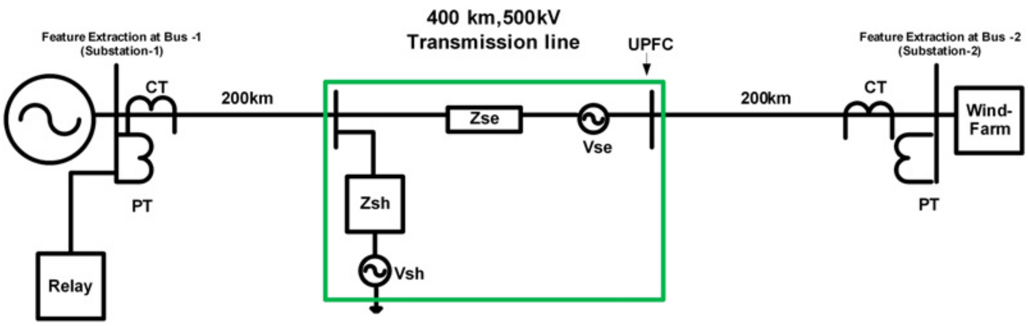


Fig. 2 *Single circuit transmission line with UPFC and wind-farm*

IET Gener. Transm. Distrib., 2014, Vol. 8, Iss. 12, pp. 2144–2152
doi: 10.1049/iet-gtd.2014.0023

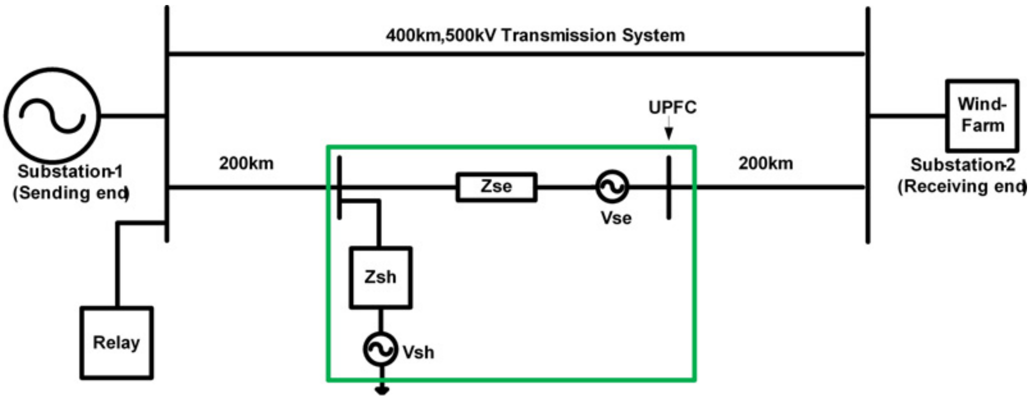


Fig. 3 *Double circuit transmission line with UPFC and wind-farm*

an AC/DC/AC IGBT-based PWM converter. The stator winding is connected directly to the 50 Hz grid, whereas the rotor is fed at variable frequency through the AC/DC/AC converter. The DFIG technology allows extracting maximum energy from the wind for low wind speeds by optimising the turbine speed, while minimising mechanical stresses on the turbine during gusts of wind. The optimum turbine speed producing maximum mechanical energy for a given wind speed is proportional to the wind speed. For wind speeds lower than 10 m/s, the rotor is running at sub synchronous speed. It runs at a hyper synchronous speed at high wind speed. Test case II is a double circuit line containing UPFC in one of the lines and wind-farm as one of the substation (shown in Fig. 3). The transmission line and other parameters remain same as given for single circuit system. Effects of mutual inductance and capacitance have been considered for this line. The details of the wind-farm modelling considered in our proposed relaying scheme are derived from [18].

Various simulation studies are carried out for the above two systems in SIMULINK environment and the DT is developed for classifying different types of faults under various operating conditions. Total number of fault cases simulated are 5 (R_f) × 4(FIA) × 10 (types of fault) × 3(fault locations) × 3 (V_{se}-UPFC voltage) × 3 (wind speed) = 5400. Along with the above faulted cases, 100 external fault cases are also simulated to check the security of the proposed relaying scheme. The resulted DT for test cases-I and II are shown in the Figs. 4 and 5, respectively. In this work, the open source data-mining software package Rattle [19] is used to develop the DT-based classification trees. The approach in Rattle to build a DT entails three steps: (i) tree growing using a learning dataset; (ii) tree pruning using cross-validation or an independent validation dataset; and (iii) selection of the optimal pruned tree. The DT growing, node splitting, tree pruning and optimal tree selection algorithms are detailed in [21]. A knowledge base comprising different types of fault scenarios is used for DT training. The 10-fold cross-validation method is used to develop the DT classification tree. Table 1 shows the effect of increase in training data set on the yield of DT and it is found that at 50–50% combination of data set, the DT starts providing 100% accuracy.

It is found from Fig. 4 that only 4 (X_{10}, X_{11}, X_{12}, X_{15}) features are taking part in DT construction for single circuit line as compared to 5 (Fig. 5) features (X_{10}, X_{11}, X_{12}, X_{15}, X_{18}) taken by double circuit line. In case of double circuit lines the mutual impedance can be as high as 50–70% of the zero-sequence impedance of the line, where as the positive- and negative-sequence coupling between the two feeders is usually less than 5–7%. Hence, it is found that in case of studied double circuit line an additional feature which is zero-sequence current difference (X_{18}) is taking part in the final decision making process along with the

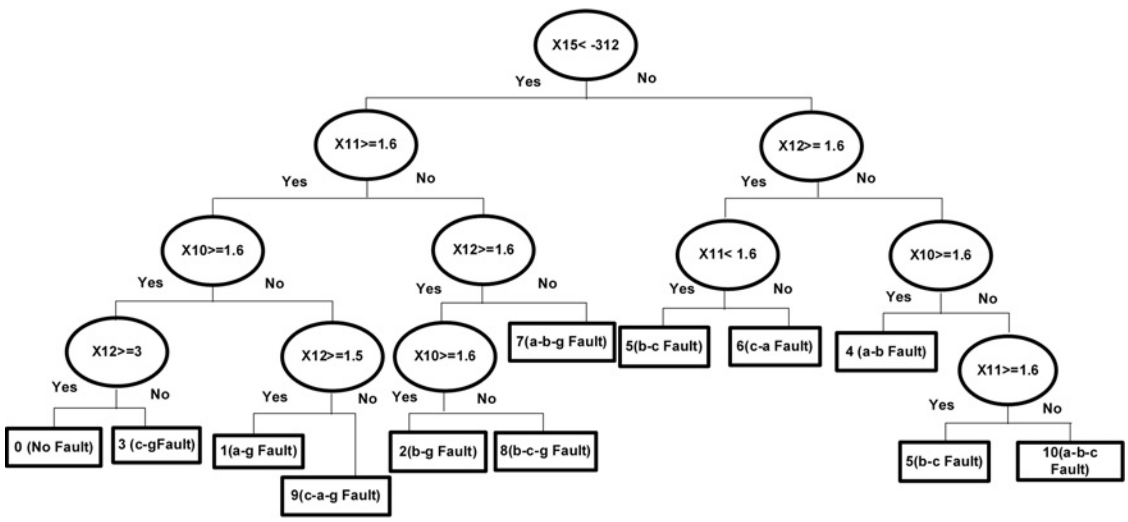


Fig. 4 *Generated DT for single circuit UPFC and wind-farm-based system*

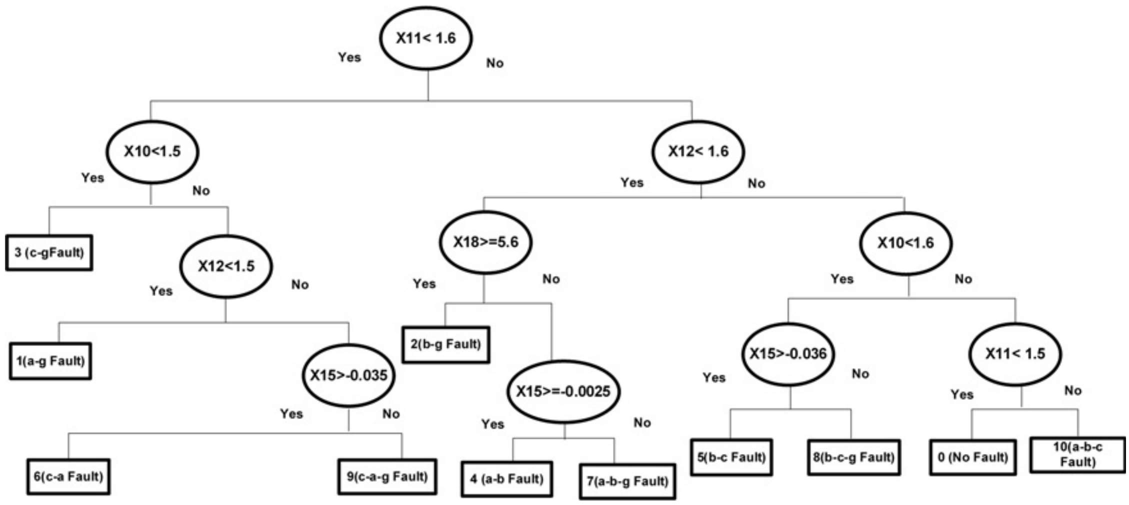


Fig. 5 *Generated DT for double circuit UPFC and wind-farm-based system*

Table 1 Accuracy comparison of DT-induced fuzzy rule-based scheme for different training and testing pattern

Training, %	Testing, %	% Overall accuracy
20	80	62
50	50	100
70	30	100

four other features which are part of the DT for single circuit line. In the DT model, each tree node is split by an input feature/variable. Variable importance is one of the key performance indexes of DT which provides information on the contribution of features to the decision making process and thus, selects the optimal numbers of features for final decision making. Thus, DT also does the task of feature selection which is used for final decision making, providing optimal performance. This is the inherent characteristics of the DT and is driven by searching all candidate predictors, and finding the split which gives the largest decrease in class impurity. The variables/feature gain credit towards their contribution by serving as primary splitters that actually split a node, or as backup splitters (surrogates) to be used when the primary splitter is missing.

The quality of the features used to classify different types of faults affects the accuracy of the classification scheme. Fig. 6 shows the correlation matrix [22] for the studied system. The degree of correlation between the features is interpreted by both the shape and the colour of the graphic elements. The shape of the graphic element shows the correlation between the candidate features as shown in Fig. 6. The colours used to shade the circles provide another clue to the strength of the correlation: the colour intensity is maximal for a perfect correlation and minimal (white) if there is no correlation. Shades of red colour are used for negative correlations and blue colour for positive correlations. It is evident from Fig. 6 that the features X_{10}, X_{11}, X_{12}, X_{15} and X_{18} are closely related to target output (fault classification). Thus, in the proposed scheme, these are the most optimal features selected for final decision making using DT.

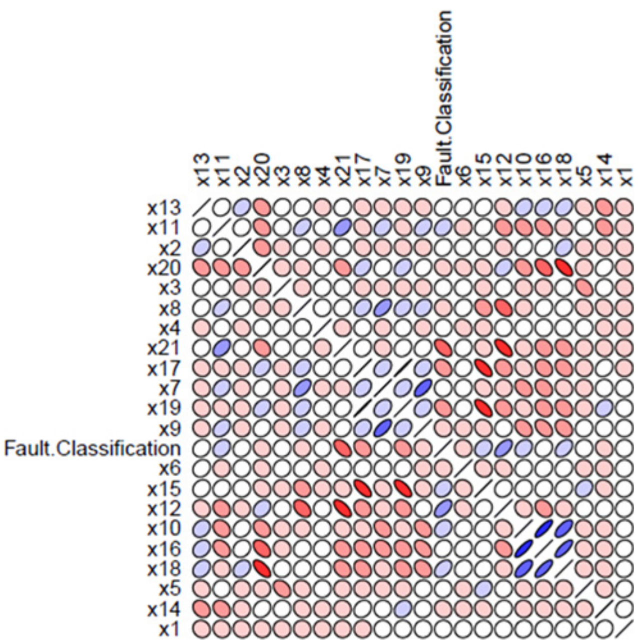


Fig. 6 *Visual summary of correlations between the 21 input features and the target output (fault classification)*

(see online version for colour)

4 Fuzzification of decision threshold

By applying fuzzy theory to individual tree branches, the partitions become a series of fuzzy regions and the sharp boundaries used by the crisp decision tree cease to exist [13, 14]. The objective of the proposed scheme is to relax these sharp decision boundaries by introducing fuzzification into the decision tree. This involves introduction of a pair of membership functions at each decision node, where each attribute is represented by a fuzzy set. The decision space is fragmented with a number of overlapping fuzzy regions [15–17]. In order to classify a particular case, all branches are fired to some degree. The degree of membership at each branch for a given attribute value depends upon the coverage of the fuzzy set for that particular branch. A fuzzy region around any decision node in a crisp decision tree is defined using a pair of complementary membership functions M_1 and M_2 around a decision threshold dt (Fig. 7). For example, a 5-node tree is represented by ten membership functions. Fig. 7 illustrates two complementary membership functions over the domain dm. dn of attribute

IET Gener. Transm. Distrib., 2014, Vol. 8, Iss. 12, pp. 2144–2152
doi: 10.1049/iet-gtd.2014.0023

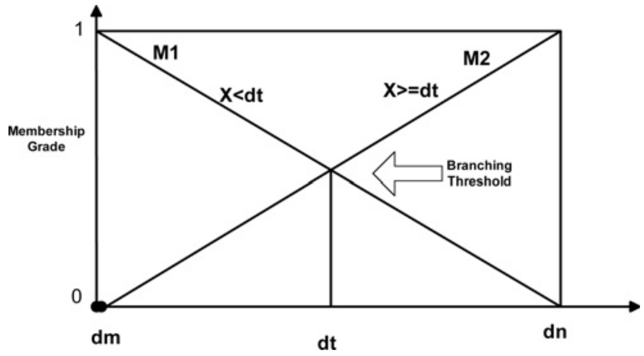


Fig. 7 *Assignment of membership grades for two complementary membership functions M_1 and M_2 over domain dm. dn for attribute i*

i, where dt_i represents the branching threshold of attribute i. A specific value x of attribute i, passing through the tree is assigned a membership grade based on its proximity to dt_i. Each membership function is having an associated domain (dm_i, dn_i) whose scope is determined by the attribute at each specific branch, which is defined as [17]

$$dm_i = dt_i - n_j\sigma_i \tag{1}$$

$$dn_i = dt_i + n_j + 1\sigma_i \tag{2}$$

where σ_i, is the standard deviation of attribute i, n is a real number, $n \to [0.0, \infty]$ used to determine the effect of σ_i on the membership function domain and dm, dn are the lower and upper bounds of membership function, respectively. In practice 'n' remains typically small, $n \to [0.0, 5]$. This is because the large value of n introduces too much fuzzification into the tree and the process of making the decision becomes too vague.

The DT is transformed to a fuzzy rule base by developing the fuzzy membership functions from the partition boundaries of the DT. From the DT boundaries, triangular MFs [17] are developed for each independent variable. For each system studied, a fuzzy interface model has been programmed in MATLAB environment. Each feature is associated with two triangular membership functions. $[B_1, B_2]$, $[C_1, C_2]$, $[D_1, D_2]$, $[E_1, E_2]$ and $[F_1, F_2]$ are the triangular membership function pairs associated with features X_{10}, X_{11}, X_{12}, X_{15} and X_{18}, respectively. All the membership functions are defined using (1) and (2). The fuzzy rule base is developed directly looking at the simple if-then-else decision logic at each and every node of the trained DT. For example 'If X_{10} is B_1 and X_{11} is C_2 and X_{12} is D_2 and X_{15} is E_1, then type of fault is $m_1(a-g$ fault)'. For single circuit line, fuzzy MFs are represented as B_1, C_2, D_2 and E_1 against the input features, X_{10}, X_{11}, X_{12} and X_{15}, respectively. The proposed algorithm does not have any output membership function. The resulting classification is a mathematical combination of the rule strengths. As a result of applying the union operator, the final grade of membership for each path within the tree is determined as a fuzzy singleton. The complete algorithm adapted in the proposed scheme for constructing a fuzzy interface framework using crisp DT is detailed in [17]. Table 1 shows the effect of increase in training data set on the yield of proposed DT-induced fuzzy rule-based scheme. It is observed that at 50–50% combination of data set, the DT starts providing 100% accuracy. 50–50% combination of dataset means out of total considered dataset, 50% of dataset are used to train the DT and rest 50% are used as testing dataset. It is also observed that the optimal features selected for 50 and 70% training data set are same which shows the efficacy of the DT for feature selection for final decision making. Thus, the 70–30% training-testing ratio is considered in the proposed study which is normally considered in the data-mining applications. The DT-induced fuzzy rule base for studied single circuit line and double circuit line are given in Tables 2 and 3, respectively, which are derived looking at the if-then-else decision logic at each and every node of the trained DT (Figs. 4 and 5).

Table 2 DT-induced fuzzy rule base for fault classification of single circuit transmission line including UPFC and wind-farm

Rules	X_{10}	X_{11}	X_{12}	X_{15}	Type of fault
R–1	B_1	C_2	D_2	E_1	1
R–2	B_2	C_1	D_2	E_1	2
R–3	B_2	C_2	0	E_1	3
R–4	B_2	C_1	D_2	E_2	4
R–5	B_2	0	D_1	E_2	5
R–6	0	C_2	D_2	E_2	6
R–7	B_1	C_1	D_2	E_1	7
R–8	0	C_1	D_1	E_1	8
R–9	B_1	C_2	D_1	E_1	9
R–10	B_1	C_1	D_1	E_2	10

Table 3 DT-induced fuzzy rule base for fault classification of double circuit transmission line including UPFC and wind-farm

Rules	X_{10}	X_{11}	X_{12}	X_{15}	X_{18}	Type of fault
R–1	B_1	C_1	D_1	0	0	1
R–2	0	C_2	D_1	0	F_2	2
R–3	B_1	C_1	0	0	0	3
R–4	0	C_2	D_1	E_2	F_1	4
R–5	B_1	C_2	D_2	E_2	0	5
R–6	B_2	C_1	D_2	E_1	0	6
R–7	0	C_2	D_1	E_1	F_1	7
R–8	B_1	C_2	D_2	E_1	0	8
R–9	B_2	C_1	D_2	E_2	0	9
R–10	B_2	C_2	D_2	0	0	10

5 Performance assessment and discussion

The previous sections deal with development of an intelligent differential relaying scheme for UPFC and wind-farm-based transmission line using DT-induced fuzzy rule-based approach. One cycle post fault differential features between both the ends of line are used to build the DT-induced fuzzy rule base for fault classification. Further, the performance assessment is carried out for different test cases considered in the proposed study. To assess the performance of the proposed relaying scheme, the statistical metrics considered are defined as follows:

(i) ***Dependability (D):*** Total number of fault cases predicted/total number of actual fault cases.
(ii) ***Yield (Y):*** Total number of correct fault cases predicted/total number of fault cases predicted.
(iii) ***Security (S)*** = Total number of external faults predicted as external fault/total number of external faults.

The dependability is one of the key indices for measuring the relay performance. This shows how effectively the relay generates the final decision and sends the tripping signals to the circuit breaker in case of the internal fault occurring on the transmission line. Yield provides the information on how accurately the faults are classified, that is, a-g fault must be classified as a-g fault and likewise. Security is a measure of assessing the performance of the proposed differential relaying scheme in case of external fault situations.

An interesting observation is made while comparing the performance indices of DT-based relaying [9] with the proposed differential relaying scheme at different fault locations of the studied single circuit transmission line. The feature set considered in [9] includes voltage and current signals of different phases. The DT of [9] and the fuzzified-DT of the proposed scheme are trained with 3850 dataset (70% of total data set considered). A total of 1650 fault cases corresponding to each fault location (10, 80 and 95% of total line length) are used as test data set. Tables 4 and 5 show the performance indices such as dependability and yield for both schemes. It is observed that the dependability and yield of the DT-based scheme substantially decreases for remote end faults compared to the proposed differential relaying scheme, which stays at 100% for both performance indices considering different fault locations. The results indicate that the use of differential features as input to DT makes the relaying scheme more reliable, compared to DT using one end features. Further, the distance relays with directional element are simulated with mho characteristics. This includes zone-1 setting to 80% of the line length with no time delay. Here, the distance protection is provided with a permissive overreach transfer trip scheme to achieve fast tripping for faults within 80–100% (fault location) range. The performance of the proposed scheme is then compared with the conventional distance relaying scheme. A total of 100 fault cases corresponding to each fault location (10, 60 and 90% of total line length) are tested ($R_F = 1\ \Omega$, FIA = 0°) and it is found that the proposed scheme is 100% dependable (Table 6) for remote end faults, where the distance relaying fails measurably with only 12%

Table 4 Dependability comparison between conventional distance relaying and proposed DT-fuzzy-based relaying for the studied single circuit line

Scheme	Dependability (fault at 10% of the line) (1650 test cases)	Dependability (fault at 80% of the line) (1650 test cases)	Dependability (fault at 95% of the line) (1650 test cases)
DT with one end data	100	40	10
DT-fuzzy	100	100	100

Table 5 Yield comparison between conventional distance relaying and proposed DT-fuzzy-based relaying for the studied single circuit line

Scheme	Yield (fault at 10% of the line) (1650 test cases)	Yield (fault at 80% of the line) (1650 test cases)	Yield (fault at 95% of the line) (1650 test cases)
DT with one end data	100	44	36
DT-fuzzy	100	100	100

Table 6 Dependability comparison between distance relaying and proposed relaying scheme for different fault locations

Scheme	*D*, % (fault at 10% of line) (100 test cases)	*D*, % (fault at 60% of line) (100 test cases)	*D*, % (fault at 90% of line) (100 test cases)
distance relaying (Ω characteristics)	100	62	12
proposed relaying scheme	100	100	100

dependability. Table 7 provides the test results of different types of faults for fault situations with and without noise for the studied single circuit line. Both the performance indices provide 100% on 50 test cases of different conditions for features without noise case, while only one miss-detection (L-L fault as LL-G) takes place for faults in noisy environment (SNR 20 dB). Thus, the proposed DT-initialised fuzzy rule-based approach is found to be accurate and robust for fault classification.

One of the important aspects of any data-mining tool is to compare its pool accuracy and testing accuracy. Classification accuracy based upon a pool of dataset is called 'pool accuracy', whereas classification accuracy based upon a particular testing dataset is called 'testing accuracy' [23]. When more and more samples are selected for the training set, both the pool accuracy and testing accuracy of fault classification show increasing trend as shown in Table 8. This fact coincides with the idea that more training samples results higher prediction ability. The most important conclusion drawn from this comparison is that the proposed scheme gives 100% testing accuracy even with fewer numbers of input data set (800 and above). Finally, the dependability and security of the proposed relaying scheme is compared with other existing schemes [20, 24] for 100 cases of crucial fault situations (fault situation with low wind penetration, UPFC bypass mode of operation, high fault resistance and so on) and 50 cases of external fault situations. The above testing is done on the single circuit transmission line and the results are depicted in Table 9. The dependability of the conventional current differential schemes [20, 24] is found to be highly sensitive to the faults with higher fault resistance and variations in wind speed. Further, the proposed scheme is found to be more secure compared to the existing differential schemes. Thus, conventional current differential schemes fail in protecting

Table 7 Dependability and yield comparison for different types of faults on testing dataset with and without noise interference for the studied single circuit line

No of cases	Actual class	Predicted class	Dependability, %	Yield, %
Features without noise				
50	L-G	L-G	100	100
50	L-L	L-L	100	100
50	L-L-G	L-L-G	100	100
50	L-L-L	L-L-L	100	100
Features with SNR 20 dB				
50	L-G	L-G	100	100
50	L-L	L-L (49 cases) + L-L-G (1 cases)	100	98
50	L-L-G	L-L-G	100	100
50	L-L-L	L-L-L	100	100

IET Gener. Transm. Distrib., 2014, Vol. 8, Iss. 12, pp. 2144–2152
doi: 10.1049/iet-gtd.2014.0023

Table 8 Comparison of pool accuracy and testing accuracy for different types of faults for the studied single circuit line

Number of input dataset	Pool accuracy, DT-fuzzy	Testing accuracy, DT-fuzzy
300	98.5	95.5
500	100	99
800	100	100
1200	100	100

Table 9 Dependability and security comparison for different types of faults between proposed scheme and other schemes

Scheme	D, % for (100 test cases)	S, % for (50 test cases)
distance relay (mho characteristics)	65	75
presented scheme [20]	72	72
presented scheme [24]	83	88
proposed scheme	100	100

Table 10 Dependability and yield comparison between only DT and proposed DT-induced fuzzy rule-based scheme for different types of faults in case of simulated double circuit line

Types of fault	Dependability of DT, %	Dependability of DT_fuzzy, %	Yield of DT, %	Yield of DT_fuzzy, %
L-G (100 test cases)	100	100	96	100
L-L (100 test cases)	100	100	92	99
L-L-G (100 test cases)	100	100	97	99
L-L-L (100 test cases)	100	100	100	100

complex transmission system such as transmission line including UPFC and wind-farm, together. However, the proposed scheme provides 100% protection measure as it is based on differential features cascaded with DT-induced fuzzy rule base for fault classification.

The proposed relaying scheme is tested for the test case-II, which is a double circuit line. To know the effect of fuzzification upon the performance indices of the relays, extensive tests are carried out for DT-based relay and the proposed DT-induced fuzzy rule-based relay considering 400 test cases. The test data set includes fault cases of both the lines, that is, transmission line with UPFC and transmission line without UPFC. Although comparing the dependability and yield for different fault situations including L-G, LL-G, L-L, LLL (100 cases from each), it is observed that (Table 10) the dependability is 100% for both the schemes, whereas the yield of the proposed DT-induced fuzzy rule-based approach is higher compared to the DT-based relay. This happens because the fuzzified-DT can handle the uncertainties of the data set better than the DT-based relaying scheme.

Another test case of a single circuit line with UPFC, wind-farm and remote end in feed (Fig. 8) is also considered for testing the proposed relaying scheme. The total line length remains same as 400 km with the UPFC at 200 km and a tap for in-feed at 370 km. The major thrust is to observe the effect of remote end in-feed upon the reliability of the proposed scheme. The dependability and yield comparisons for different fault situations including L-G, LL-G, L-L, LLL (100 cases from each) are shown in Table 11 and it is observed that both the performance indices achieve 100% accuracy. The reason being that the fault becomes severe when contributed to from the in-feed in case of in-feed from remote end. Further, it is observed that the fuzzy rule-bases developed for single circuit and double circuit lines are almost the same, except inclusion of one feature (X_{18}) in case of double circuit line compared to single circuit line. This happens as double circuit line is more complex, including mutual coupling effect which is absent in case of single circuit lines. Moreover, even if the topology of power system changes, the proposed DT-fuzzy approach is able to provide a comprehensive solution for effective and reliable protection decision.

The computational time of the proposed DT-induced fuzzy rule-based scheme is 0.06 s for the test data set (1620 data-30% of 5400) on a PC (Core(TM)i5–2400 CPU@3.10 GHz). This computational time is the time consumed by fuzzified-DT to provide final relaying decision on the test data set. However, when individual fault case is considered, the response time is 25 ms (1 and 1/4th cycle) from fault inception, which includes 20 ms time for DFT-based pre-processing and 5 ms for DT-induced fuzzy rule-based classification. Thus, the speed of the proposed scheme is suitable for relaying application. The performance of the proposed relaying scheme is validated assuming the communication platform is established. The response time of the proposed relay is 25 ms, excluding the communication delay, and inclusion of communication delay will not impact the performance indices such as dependability and security of the relay. However, the response time of the relay will be delayed by the

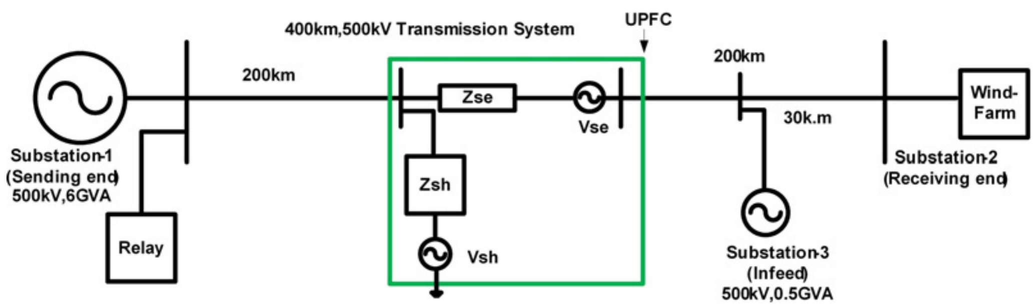


Fig. 8 *Schematic representation of studied in-feed line*

Table 11 Dependability and yield comparison for different types of faults in case of simulated in-feed circuit

Types of fault	Dependability of DT_fuzzy, %	Yield of DT_fuzzy, %
L-G (100 test cases)	100	100
L-L (100 test cases)	100	100
L-L-G (100 test cases)	100	100
L-L-L (100 test cases)	100	100

communication delay time (which is in the order of 4 ms for long transmission lines [25, 26]).

6 Conclusions

The paper presents a DT-induced fuzzy rule-based differential relaying scheme for transmission line employing UPFC and wind-farm together. The process starts at pre-processing the voltage and current signal using DFT-based pre-processor. The derived differential features are used to build the optimal DT-induced fuzzy rule base for selecting the faulty phase(s) involved in the fault process. The proposed scheme is extensively tested for single circuit and double circuit transmission lines. The results obtained with respect to performance indices such as dependability, yield and security show the effectiveness of the proposed relaying scheme. The speed response of the proposed relaying scheme falls within 1 and 1/4 cycles from the fault inception, which is fast enough for relaying purpose. Thus, the proposed DT-induced fuzzy rule-based differential relaying scheme is reliable and robust for transmission line protection considering FACTs and wind-farms together, which are embedded part of modern Power system.

7 Acknowledgment

This research work is supported by Department of Science and Technology, Government of India, (SR/FTP/ETA-78/2010).

8 References

1 Hingorani, N.G., Gyugyi, L.: 'Understanding FACTS Concepts and Technology of Flexible AC Transmission Systems' (IEEE Press, New York, 2000)
2 Samantaray, S.R.: 'Decision tree-based fault zone identification and fault classification in flexible AC transmissions-based transmission line', *IET Gener. Transm. Distrib.*, 2009, **3**, (5), pp. 425–436
3 Pradhan, A.K., Geza, J.: '*Adaptive distance relay setting for lines connecting wind-farms*', *IEEE Trans. Energy Convers.*, 2007, **22**, (1), pp. 206–213
4 Jang, S.I., Choi, J.H., Kim, J.W., Choi, D.M.: 'An adaptive relaying for the protection of a Wind-farm interconnected with distribution networks'. Proc. IEEE PES Transmission Distribution Conf. Expo., 7–12 September 2003, vol. 1, pp. 296–302
5 Yang, G., Dong, M., Zhou, Z., *et al.*: 'The influences and countermeasures of wind-farm access to transmission line differential protection'. IEEE Int. Conf. Power Electronics and Machines in Wind Applications (PEMWA), 16–18 July 2012, pp. 1–4
6 Bin, L., Meng, L.: 'Analysis of directional pilot protection on transmission line with wind-farm integration'. IEEE Int. Conf. Advanced System Automation and Protection (APAP), 16–20 October 2011, pp. 303–308
7 Wang, X., Zhou, H., Aggarwal, R.K., Beaumont, P.: 'Performance evaluation of a distance relay as applied to a transmission system with UPFC', *IEEE Trans. Power Deliv.*, 2006, **21**, (3), pp. 1137–1147
8 Dash, P.K., Samantaray, S.R., Panda, G.: 'Fault classification and section identification of an advanced series-compensated transmission line using support vector machine', *IEEE Trans. Power Deliv.*, 2007, **22**, (1), pp. 67–73
9 Jamehbozorg, A., Shahrtash, S.M.: 'A decision tree-based method for fault classification in single-circuit transmission lines', *IEEE Trans. Power Deliv.*, 2010, **25**, (4), pp. 2190–2196
10 Kamwa, I., Samantaray, S.R., Joos, G.: 'Development of rule-based classifiers for rapid stability assessment of wide-area post-disturbance records', *IEEE Trans. Power Syst.*, 2009, **24**, (1), pp. 258–270
11 Jena, M.K., Samantaray, S.R.: 'Real time digital simulator implementation of data-mining-based integrated impedance relay for compensated lines', *Int. Trans. Electr. Energy Syst.*, 2014. doi: 10.1002/etep.1920
12 Banfield, R.E., Hall, L.O., Bowyer, K.W., Kegelmeyer, W.P.: 'A comparison of decision tree ensemble creation techniques', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2007, **29**, (1), pp. 173–80
13 Yuan, Y., Shaw, M.J.: 'Induction of fuzzy decision trees', *Fuzzy Sets Syst.*, 1995, **69**, pp. 125–139
14 Zeidler, J., Schlosser, M.: 'Continuous-valued attributes in fuzzy decision trees'. Proc. of Information Processing and Management of Uncertainty in Knowledge-Based Systems, Granada, 1996, pp. 395–400
15 Tsang, E.C.C., Wang, X.Z., Yeung, Y.S.: 'Improving learning accuracy of fuzzy decision trees by hybrid neural networks', *IEEE Trans. Fuzzy Syst.*, 2000, **8**, (5), pp. 601–614
16 Wang, X.Z., Chen, B., Qian, G., Ye, F.: 'On the optimization of fuzzy decision trees', *Fuzzy Sets Syst.*, 2000, **112**, pp. 117–125
17 Crockett, K., Bandar, Z., Mclean, D., O'Shea, J.: 'On constructing a fuzzy inference framework using crisp decision trees', *Fuzzy Sets Syst.*, 2006, **157**, (1), pp. 2809–2832
18 Ekanayake, J.B., Holdsworth, L., Wu, X., Jenkins, N.: 'Dynamic modeling of doubly fed induction generator wind turbines', *IEEE Trans. Power Syst.*, 2003, **18**, (2), pp. 803–809
19 Rattle (the R Analytical Tool to Learn Easily), by D.Williams: ver. 2.3, May 2008. [Online]. Available: http://rattle.togaware.com/
20 Xu, Z.Y., Du, Z.Q., Ran, L., Yang, Y.K., He, J.L.: 'A current differential relay for a 1000-kV UHV transmission line', *IEEE Trans. Power Deliv.*, 2007, **22**, (3), pp. 1392–1399
21 Breiman, L., Friedman, J., Olshen, R.A., Stone, C.J.: 'Classification and regression trees' (Wadsworth, Pacific Grove, CA, 1984)
22 Kamwa, I., Samantaray, S.R., Joos, G.: 'Catastrophe predictors from ensemble decision-tree learning of wide-area severity indices', *IEEE Trans. Smart Grid*, 2010, **1**, (2), pp. 148–158
23 Xi-Zhao, W., Dong, L.C., Yan, J.H.: 'Maximum ambiguity-based sample selection in fuzzy decision tree induction', *IEEE Trans. Knowl. Data Eng.*, 2012, **24**, (8), pp. 1491–1505
24 Dambhare, S., Soman, S.A., Chandorkar, M.C.: 'Adaptive current differential protection schemes for transmission-line protection', *IEEE Trans. Power Deliv.*, 2009, **24**, (4), pp. 1832–1841
25 Michel, G., Pleinka, G., Adamiak, M., *et al.*: Digital communications for relay protection' [Online]. Available: http://www.pes-psrc.org/
26 Perera, N., Rajapakse, A.D.: 'Series-compensated double-circuit transmission line protection using directions of current transients', *IEEE Trans. Power Deliv.*, 2013, **28**, (3), pp. 1566–1575

doi: 10.1049/iet-gtd.2014.0023