A Data-Mining Model for Protection of FACTS-Based Transmission Line

S. R. Samantaray, *Senior Member, IEEE*

***Abstract*—This paper presents a data-mining model for fault-zone identification of a flexible ac transmission systems (FACTS)-based transmission line including a thyristor-controlled series compensator (TCSC) and unified power-flow controller (UPFC), using ensemble decision trees. Given the randomness in the ensemble of decision trees stacked inside the random forests model, it provides effective decision on fault-zone identification. Half-cycle postfault current and voltage samples from the fault inception are used as an input vector against target output "1" for the fault after TCSC/UPFC and "−1" for the fault before TCSC/UPFC for fault-zone identification. The algorithm is tested on simulated fault data with wide variations in operating parameters of the power system network, including noisy environment providing a reliability measure of 99% with faster response time (3/4th cycle from fault inception). The results of the presented approach using the RF model indicate reliable identification of the fault zone in FACTS-based transmission lines.**

***Index Terms*—Distance relaying, fault-zone identification, random forests (RFs), support vector machine (SVM), thyristor-controlled series compensator (TCSC), unified power-flow controller (UPFC).**

I. Introduction

INCREASED demand of bulk power transfer in the modern power network has led to an increased focus on transmission constraints and alleviation. Flexible ac transmission systems (FACTS) [1] devices offer a versatile alternative to conventional reinforcement methods. Among them, the thyristor-controlled series compensator (TCSC) [2] and unified power-flow controller (UPFC) [3] are important FACTS devices, which are used extensively for improving the utilization of the existing transmission system. The presence of TCSC in fault loop not only affects the steady-state components but also the transient components. The controllable reactance, the metal–oxide varistors (MOVs) protecting the capacitors, and the air-gap operation make the protection decision more complex and, therefore, the conventional relaying scheme based on fixed settings finds limitations. On the other hand, UPFC offers new horizons in terms of power system control. While the use of UPFC improves the power transfer capability and stability of a power system, certain other problems emerge in transmission-line protection [4]–[6], greatly affecting the reach of the distance relay.

In the FACTS-based transmission line, if the fault does not include FACTS device, then the impedance calculation is like an ordinary transmission line, and when the fault includes FACTS, then the impedance calculation accounts for the impedances introduced by FACTS device. The line impedance is compared with the protective zone and if the line impedance is less than the relay setting, then the relay issues a signal to trip the circuit breaker (CB). Further, for similar types of faults, the current level may be of the same order at two different points of the transmission line, (before and after TCSC/UPFC). Thus, before the apparent impedance to the fault point is computed, a more reliable and accurate fault-zone identification technique is necessary for safe and reliable operation of the distance relay. The correct fault-zone identification in presence of the FACTS devices, such as TCSC and UPFC in the transmission line, is one of the critical tasks to be dealt with.

Recent techniques based on neural networks [7], [8], find limitations, since they require a large number of neurons to model the structure of the network involving large training sets and training time. Recently, a hybrid technique using a wavelet transform combined with support vector machine (SVM) [9], [10] has been proposed for fault-zone identification in the TCSC line. The aforementioned work finds limitations since the wavelet transform is highly prone to noise and provides erroneous results even with a signal-to-noise ratio (SNR) of 30 dB [12]. Also, the computational time of SVM is higher compared to the proposed ensemble DTs-based data-mining model, which puts constraints on the online realization of SVM-based relays for distance relaying applications, where speed and accuracy are prime considerations. Thus, there is a strong motivation to build up an accurate and faster data-mining model for fault-zone identification in FACTS-based transmission lines.

The proposed research is based on a data-mining model known as ensemble decision trees [13]–[18], also known as random forests (RFs), for fault-zone assessment in a FACTS (TCSC/UPFC)-based transmission line. Half-cycle postfault current and voltage samples (time-domain data samples) are used as inputs to the RF against target outputs "−1" for faults before TCSC/UPFC and "1" for faults after TCSC/UPFC. The RF is trained to build a data-mining model with an extensive data set derived from a series of fault simulations. The proposed technique is tested on wide variations in operating parameters in the power system network, including a noisy environment and, was found to be accurate and robust for fault-zone identification in TCSC/UPFC-based transmission lines. The following sections deal with RFs, systems studied, results, analysis, discussion, and conclusions.

Manuscript received July 20, 2011; revised May 28, 2012, October 13, 2012, and December 12, 2012; accepted January 18, 2013. Date of publication February 22, 2013; date of current version March 21, 2013. This work was supported by the Department of Science and Technology, Government of India (SR/FTP/ETA-78/2010). Paper no. TPWRD-00614-2011.

The author is with the School of Electrical Sciences, Indian Institute of Technology Bhubaneswar, Orissa 751013, India (e-mail: sbh_samant@yahoo.co.in).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRD.2013.2242205

II. Random Forests

RFs are a large combination of de-correlated tree predictors such that each tree depends on the values of a random vector sampled independently. Individual trees are noisy and unstable, but since when grown sufficiently deep, they have relatively low bias. Therefore, they are ideal candidates for ensemble growing since they can capture complex interactions, while fully benefitting from aggregation-based variance reduction. Using a random selection of features to split each node and resampling (with replacement) the training set to grow each tree yields error rates that are de-correlated and more robust with respect to noise. The generalization error of forests converges to a limit since the number of trees in the forest increases.

The basic idea of most ensemble tree growing procedures is that for the kth tree ($k \leq n_{\mathrm{tree}}$, the number of trees in the ensemble) a random vector Φ_k is generated, independent of the past random vectors $\Phi_k, \ldots \Phi_{k-1}$, but with the same distribution, and a single tree is grown using the training set S and the set of attributes in Φ_k, resulting in a classifier $T_k(x, \Phi_k)$ where x is an input vector. In random split selection, Φ consists of a number n_{try} of independent random integers where $n_{\mathrm{try}} < n_a$ is the number of attributes in.

An RF consists of a collection of tree-structured classifiers $\{T_k(x, \Phi_k), k = 1, \ldots, n_{\mathrm{tree}}\}$, where $\{\Phi_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x. An algorithmic view of the RF growing process is summarized as follows [13]:

1) For $k = 1$ to n_{tree}:
   a) Draw a boostrap sample S^* of size N from the training data S
   b) Grow a random forest tree $T_k(x, \Phi_k)$ to the boostrapped data, by recursively repeating the steps below for each terminal noted of the tree, until the no other split is possible (unpurned tree of maximal depth):
      i) Select n_{try} variables from the n_a features
      ii) Pick the best variable/split-point among the n_{try}
      iii) Split the node into two daughter nodes
2) Output the ensemble of trees $\{T_k(x, \Phi_k), k = 1, \ldots, n_{\mathrm{tree}}\}$.

A traditional decision tree essentially represents an explicit decision boundary, and an instance E is classified into class c if E falls into the decision area (a leaf in the decision tree) corresponding to c [16]. The class probability p(c|E) is typically estimated by the fraction of instances of class c in the leaf into which E falls. This probability estimate is very crude when the tree is pruned because all of the instances falling into the same leaf have the same class probability. More accurate probability estimates require unpruned trees [19], which are the backbone of the RFs. Stated otherwise, the RF predictor has the additional advantage of providing a stability or instability level of the event through probability-based ranking. Assuming that the probability estimates from individual trees are random variables, each with variance σ^2, the average variance is $\sigma^2/n_{\mathrm{tree}}$, which confirms that the RF leads seamlessly to improved probability estimates [13].

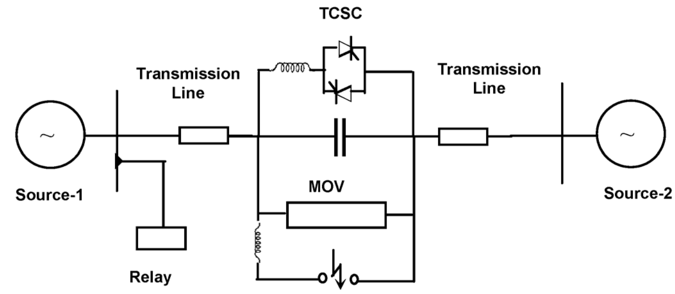

Fig. 1. Transmission line with TCSC.

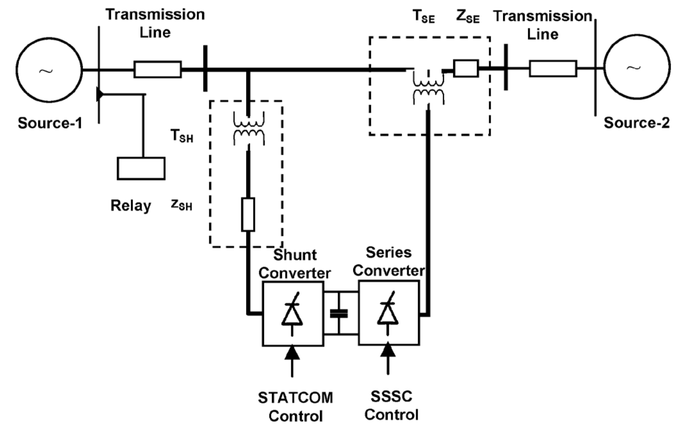

Fig. 2. Transmission line with UPFC.

Although the RF is a relatively young data-mining tool, people [20]–[22] have started recognizing its strengths: 1) it is simple and easy to use; 2) very high accuracy; 3) its relatively robust to outliers and noise; 4) it gives useful internal estimates of error, strength, and correlation; 5) not overfitting if selecting a large number of trees; and 6) insensitive to the choice of split.

III. System Studied

A. TCSC- and UPFC-Based Line

A 400-kV, 50-Hz power system is illustrated in Fig. 1. The power system consists of two sources: TCSC [10] and associated components, and a 300-km transmission line. The transmission line has zero-sequence impedance $Z_0 = 105.65 + j356.71\ \mathrm{ohm}$ and positive-sequence impedance $Z_1 = 10.52 + j115.45\ \mathrm{ohm}$. $E_s = 400$ kV and $E_R = 400\angle\delta$ kV.

The TCSC is designed such that it provides 30% compensation at 180° (minimum) and 40% compensation at 150° (maximum) firing angle, and in this study, the firing angle is varied within this range. The TCSC is placed at 30%, 50% and 80% of the transmission line to assess the impact of TCSC placement on the performance of the developed data-mining model. Similarly, UPFC [11] is placed at 30%, 50%, and 80% of the line (Fig. 2) with variations in series injected voltage and phase angle. The simulation includes all ten types of shunt faults (L-G: Line-Ground, LL-G: Line-Line-Ground, LL: Line-Line, and LLL: Line-Line-Line) with different fault