Enhancing the Protection scheme for FACTS-Based Transmission Lines using a Data Mining Model

Pallavi P. Wankhede[1], Rajesh K. Nasare[2]

Department of Information Technology, Yeshwantrao Chavan College of engineering, Nagpur, India

Abstract: ***This review paper is based on the use of the thyristor control series compensator (TCSC) for the fault zone identification and protection of FACTS based transmission lines by using data mining model with the use of ensemble decision trees. In this work we are comparing the existing data mining technologies which are available for the fault zone in the transmission line. The kd tree algorithm is used to detect the fault zone and the random forest algorithm is applied to that zone for enhancing the prior results of the random forest which is only applied to data.The random forest algorithm gives larger accuracy and reliability as compared to the other existing technologies .***

Keywords: Flexible ac transmission system (FACTS),KD(K-dimensional) tree, Random Forest algorithm (RF), Thyristor control series compensator (TCSC)

1. Introduction

Transmission lines are very much prone to failures and faults. It is very difficult to fix the fault manually when it occurs. There are many methods available in the past in order for detection of fault diagnosing application .The growth of available data in the electric power industry motivates the adoption of data mining techniques. The companies in this scope faces several difficulties that they could not benefit themselves from data mining approach. One of the reasons is that mining power systems data is an Interdisciplinary task. Typically, electrical and computer engineers or scientists need to work together in order to achieve breakthroughs, interfacing power systems and data mining at a mature level of cooperation. Another reason is the lack of freely available and standardized benchmarks. Because of that, most previous research in this area used proprietary datasets, which makes difficult to compare algorithms and reproduce results.. In the first part gives the brief overview on how the existing data mining techniques used in power systems. There are several works which had been done to introduce data mining techniques to people with background knowledge of power systems. In contrast, this work assumes previous knowledge of data mining illustrate some fundamental concepts of power systems and explains the kind of problems that transmission lines tries to solve with data mining.

Here, the well-known Alternative Transients Program simulator (ATP) is used to create comprehensive labeled dataset. Such datasets are key components which allow reproducing research results in different locations, which are important given the large number of parameters to be tuned in a fault classification system. Therefore, in order to encourage reproducible research, this work also provides detailed information about the adopted parameters. In next part a brief description of data mining applications in power systems is provided and their comparison is made on the basis of various problems in the transmission lines. Figure 1 illustrated the TCSC transmission line[1].

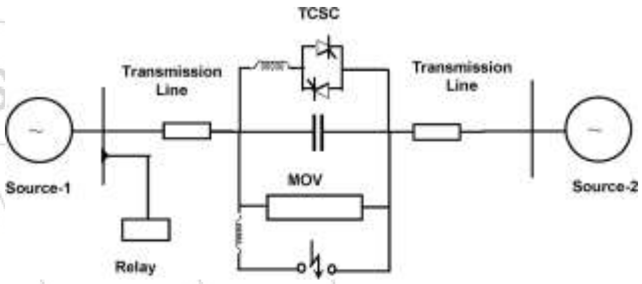

Figure 1: Transmission line with TCSC

2. Existing Work

There are number of authors who proposed methods for providing protection to the flexible transmission lines like support vector machine (SVM)[4], Artificial Neural Network (ANN)[8][9] and fuzzy neuro[11] system many more . However, each of them has their own set of disadvantages. Following table describes Summary of techniques and problems of data mining in power systems.

Table 1: Summary of techniques and problems of existing techniques

Reference	Technique	Problems
Vale,2008	Decision tree	Characterization and classification of consumer
Taghizadeh,2007	Neural network	Faults classification and location
Gouveia,2005	Decision tree	Electric energy consumer
Brito,2006	Neural Network	Faults detection and classification
Choudhury,2005	Decision tree and neural network	Faults classification
Yung,2004	Statistical analysis	Detection of substations most sensible to disturbances
Panda,2007	Support vector machine	Classification and identification of series compensator
Kerunovic,2002	Fuzzy/Neural netwok	Fault classification
Gulbagriyanik,2002	Neural network	Detection and diagnosis of transient and faults
Keerthipala,1998	Neural Network	Fault classification

3. Methodology

A. K-d Tree

A k-d tree, or k-dimensional tree, is a data structure used for organizing some number of points in a space with k dimensions. It is a binary search tree with other constraints imposed on it. K-d trees are very useful for range and nearest neighbour searches. The root-cell of this tree represents the entire simulation volume. The other cells represent rectangular sub-volumes that contain the mass, center-of-mass, and qudrupole moment of their enclosed regions. It was one of the early structures used for indexing in multiple dimensions. Each level of K-d tree partitions the space into two partitions, the partitioning is done along one dimension of the node at the top level of the tree, along another dimension in nodes at the next level, and so on, iterating through the dimensions. The partitioning proceeds in such a way that, at each node, approximately one half of the points stored in the subtree fall on one side, and one half fall on the other. Partitioning stops when a node has less than a given maximum number of points [12].Searching for a nearest neighbor in a *k*-d tree
proceeds as follows:

1) Starting with the root node, the algorithm moves down the tree recursively, in the same way that it would if the search point were being inserted (i.e. it goes left or right depending on whether the point is less than or greater than the current node in the split dimension).
2) Once the algorithm reaches a leaf node, it saves that node point as the "current best".
3) The algorithm unwinds the recursion of the tree, performing the following steps at each node:

   I. If the current node is closer than the current best, then it becomes the current best.
   II. The algorithm checks whether there could be any points on the other side of the splitting plane that are closer to the search point than the current best. In concept, this is done by intersecting the splitting hyper-plane with a hyper-sphere around the search point that has a radius equal to the current nearest distance. Since the hyperplanes are all axis-aligned this is implemented as a simple comparison to see whether the difference between the splitting coordinate of the search point and current node is less than the distance (overall coordinates) from the search point to the current best.
   III. If the hyper-sphere crosses the plane, there could be nearer points on the other side of the plane, so the algorithm must move down the other branch of the tree from the current node looking for closer points, following the same recursive process as the entire search.
   IV. If the hyper-sphere doesn't intersect the splitting plane, then the algorithm continues walking up the tree, and the entire branch on the other side of that node is eliminated. When the algorithm finishes this process for the root node, then the search is complete.

B. Random Forest

Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The generalization error for forests converges a.s. to a limit as the number of trees in the forest becomes large. The generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them. Using a random selection of features to split each node yields error rates that compare favorably to Adaboost (Freund and Schapire[1996]), but are more robust with respect to noise. Internal estimates monitor error, strength, and correlation and these are used to show the response to increasing the number of features used in the splitting. Internal estimates are also used to measure variable importance. These ideas are also applicable to regression.

C. Random forest Algorithm

The random forests algorithm is as follows [1]:
1) Draw ntree bootstrap samples from the original data.
2) For each of the bootstrap samples, grow an unpruned classification or regression tree, with the following modification: at each node, rather than choosing the best split among all predictors, randomly sample m try of the predictors and choose the best split from among those variables. (Bagging can be thought of as the special case of random forests obtained whenm try = p, the number of predictors.)
3) Predict new data by aggregating the predictions of the ntree trees (i.e., majority votes for classification, average for regression). An estimate of the error rate can be obtained, based on the training data, by the following:
   a) At each bootstrap iteration, predict the data not in the bootstrap sample (what Breiman calls "out-of-bag", or OOB, data) using the tree grown with the bootstrap sample.
   b) Aggregate the OOB predictions. (On the average, each data point would be out-of-bag around 36% of the times, so aggregate these predictions.) Calculate the error rate, and call it the OOB estimate of error rate.

4. Proposed Work

In this paper, the Random Forest (RF) algorithm for the diagnosis of fault is discussed. RF is good generalized than any other fault diagnos is applications. The main aim is the classification of the type of fault in the transmission lines. In the further work sampling of current and voltage is carried out, calculated and then they utilized as an optimal learning pattern. The k- dimensional tree algorithm is first applied to the generated data sets. To find the accurate region for the affected area the range query algorithm is applied and results are generated. Then the RF is applied to results for finding the more accurate results. By using the RF method simulations show that the identification of each class of fault is accurately done as compared to previously existing methodologies as; Support Vector Machine,Fuzzy Theory, ArtificialNeuralNetwork, ExpertSystem, PetriNet,. The result of simulation demonstrates the effectiveness of proposed method to the identification of faults.

The data-mining model used is also known as ensemble decision trees as random forests (RF), which is used for detecting the fault-zone in a FACTS based (TCSC) transmission line. The simulated data from the simulation model are used as inputs to the RF against target outputs. The RF is provided with the training to build a data-mining model with an broad data set developed from a series of fault simulations. The various operating system parameters are used for testing in the power system network, considering noisy surrounding. They are very accurate and robust for fault-zone identification in TCSC-based transmission lines. The simulations in this work relied on Matlab , which has many learning algorithms. Specifically, the work used decision trees multilayer artificial neural network (ANN) trained with back propagation. There are number of transmission parameters are use to generate the simulated data base depending on the location , fault type, firing angles, fault resistances etc. They are as follows[1]: The simulation model is design by using the MATLAB Simulink toolbox. Total simulations carried for TCSC line are 5 (Rf) × 3 (Zs) × 4 (FIA) × 10 (types of faults)×2(reverse power flow) × 2 (load changes) × 4 (firing angels) ×4(locations)=38400.

Figure 2 shows some amount of the data sets generated. Column A denotes the sample voltage and B denotes the sample current and C denotes the location. Figure 3 is the simulation model from which the data sets are generated.

	A	B	C
1	0.417022	0.087482	1
2	0.720324	0.22731	1
3	0.000114	0.314377	1
4	0.302333	0.174766	1
5	0.146756	0.607094	1
6	0.092339	0.413586	1
7	0.18626	0.816352	1
8	0.345561	0.18513	1
9	0.396767	0.701877	1
10	0.538817	0.240356	1
11	0.419195	0.574219	1
12	0.68522	0.348988	1
13	0.204452	0.056964	1
14	0.878117	0.228814	1
15	0.027388	0.664103	1
16	0.670468	0.49725	1
17	0.417305	0.519016	1
18	0.55869	0.17472	1
19	0.140387	0.570716	1
20	0.198101	0.996753	1
21	0.800745	0.816835	1
22	0.968262	0.594373	1
23	0.313424	0.975989	1
24	0.692323	0.901563	1
25	0.876389	0.595608	1

Figure 2: Snapshot of data generated

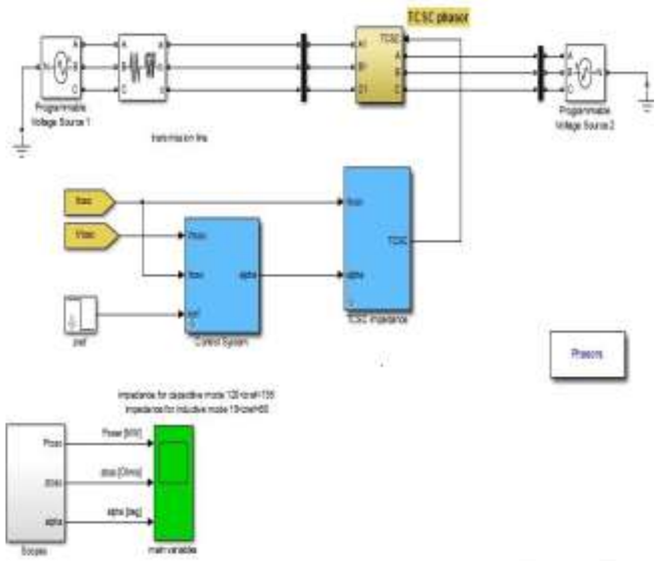

Figure 3: Simulation model using TCSC

5. Experimental Results

The following are the results which are calculate for detecting the fault zone in the transmission line by first applying the kd tree algorithm then fiding the range by range query algorithm followed by the random forest algorithm.

Figure 5: GUI window

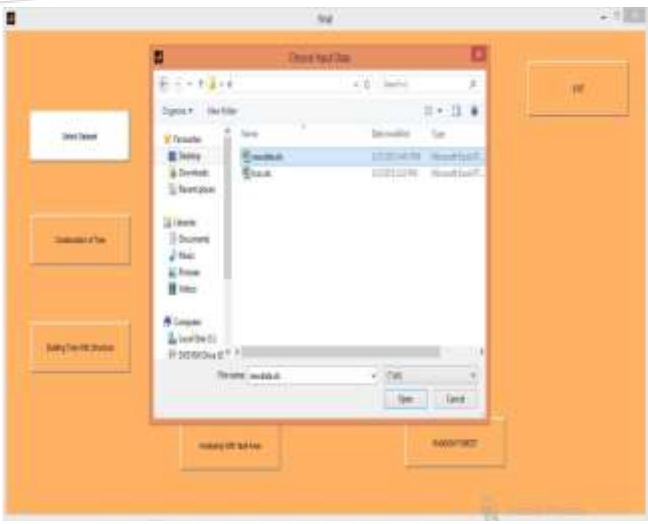

Figure 4: Selection of Data base

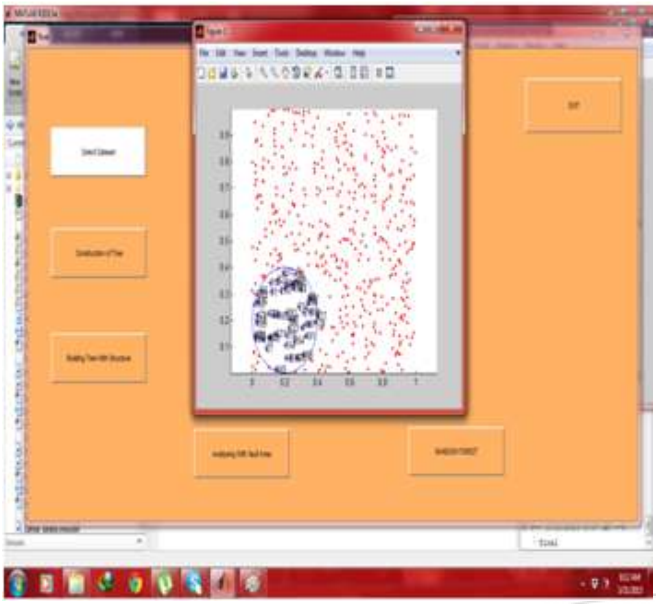

Figure 6: Construction of tree

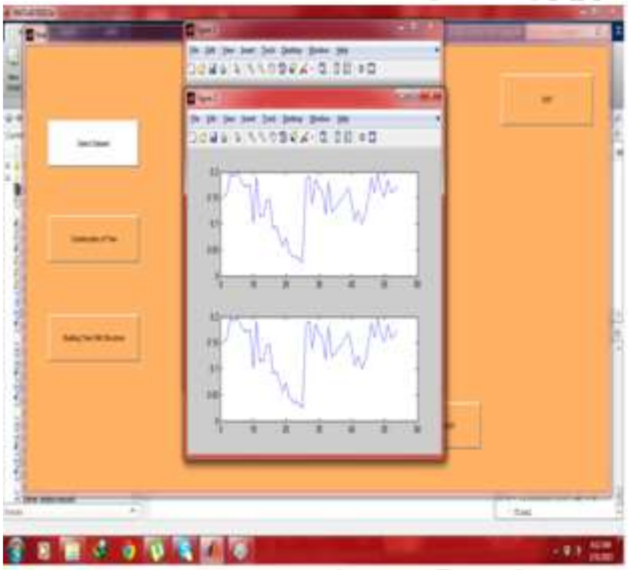

Figure 7: Building of a tree structure

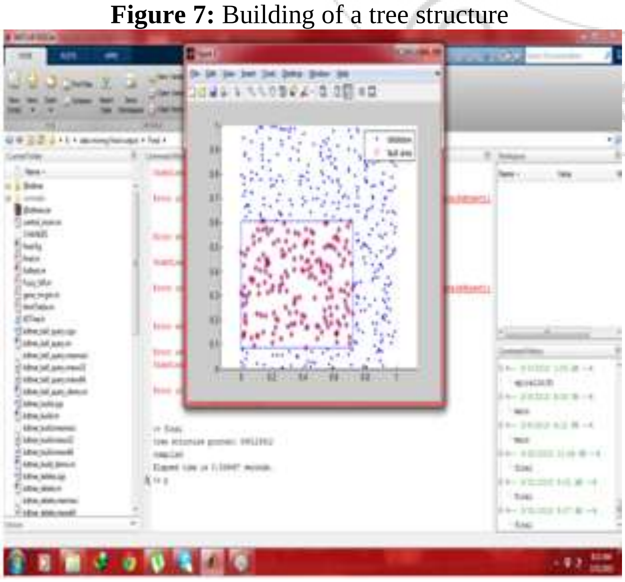

Figure 8: Diagnosis of fault area

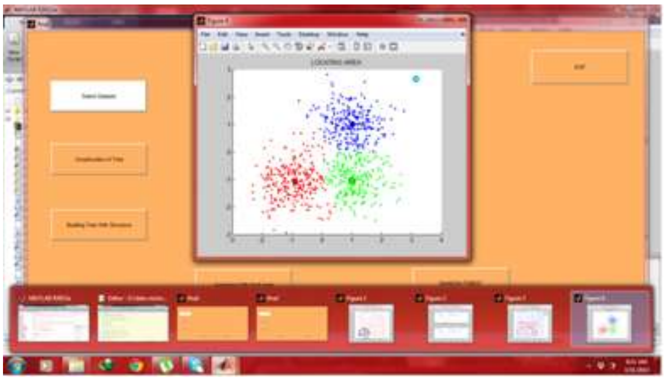

Figure 9: Locating the fault area

6. Conclusion and Future Scope

Here in this paper the proposed technique provides the data mining technique for the process that to the tree devices of RF modules. To identify the fault detection of the process based on the FACTS devices and the transmission of its based lanes with an accreted data readings and the reliability of process has more than 99% RF trees for this we are maintaining a way of data mining algorithm for the detection process through this we has found the ¾ of the results has come on success to identify the transmission lane power voltage. In the existing system they had not supported the fault detection for the process and the values just the secondary data accumulations of supply only and for that we proposed simulation model to generate data sets and then to identify the fault values and to modify it. Here in this the further process is the application of random forest to the results and then its reliable identifier will compare the data of reading and to that modification we can simply display and we can find out the normal results for the process. In future work, there are more investigations that can be performed among the different sampling times needed to ensure best performance thereby generating best results.

References

[1] S. R. Samantaray, *"A Data Mining Model for Protection of FACTS-Based Transmission Line*",Feb2013

[2] *D. S. Siroky*, "Navigating random forests and related advances in algorithmic modeling," Stat. Rev. vol. 3, pp. 147–163,2009.[Online].Available: http://projecteuclid.org/DPubS?service=UI&version=1.0&verb=Display&handle=euclid.ssu

[3] M. Khederzadeh and T. S. Sidhu, *"Impact of TCSC on the protection of transmission lines," IEEE Trans. Power Del.*, vol. 21, no. 1, pp. 80–87, Jan. 2006.

[4] P. K. Dash, S. R. Samantaray, and G. Panda, *"Fault classification and section identification of an advanced series-compensated transmission line using support vector machine,"* IEEE Trans. Power Del., vol. 22. no. 1, pp. 67–73, Jan. 2007

[5] Paserba, J.J.; Miller, N.W.; Larsen, E.V.; Piwko, R.J., "*A thyristor controlled series compensation model for power system stability analysis,"* Power Delivery, IEEE Transactions on , vol.10, no.3 Jul 1995

[6] H. Wang, W. W. L. Keerthipala, *"Fuzzy-neuro approach to fault classification for transmission line*

protection," IEEE Trans. Power Del. 13 (4) (1998) 1093-1104.

[7] Monedero, I.; Leon, C.; Ropero, J.; Garcia, A.; Elena, J.M.; Montano, J. C. (2007)."*Classification of Electrical Disturbances in Real Time Using Neural Networks".* IEEE Transactions on Power Delivery, Vol. 22, No. 3, (July 2007) 1288 – 1296, ISSN 0885-8977.

[8] Vasilic, S.; Kezunovic, M. (2005*) "Fuzzy ART Neural Network Algorithm for Classifying the Power System Faults"*. IEEE Transactions on Power Delivery, Vol. 20, No. 2, (April2005) 1306-1314, ISSN 0885-8977.

[9] Vasilic, S.; Kezunovic, M. (2002)."*An Improved Neural Network Algorithm for Classifying the Transmission Line Faults"*. IEEE Power Engineering Society Winter Meeting, pp. 918 – 923. ISBN 0-7803-7322-7, Jan 2002.

[10] Kezunovic, M. ;Vasilic, S.; Gul-Bagriyanik, F. (2002). "*Advanced Approaches for Detecting and Diagnosing Transients and Faults*". 2002.

[11] Huisheng Wang; Keerthipala, W.W.L. (1998).*"Fuzzy-neuro approach to fault classification for transmission line protection*".IEEE Transactions Power Delivery, Vol.13, No. 4, 1093-1104, ISSN 0885-8977.

[12] William R. Mark, Gordon Stoll, (2006) "*Fast kd-tree Construction with an Adaptive Error-Bounded Heuristic"*, Warren Hunt, IEEE Symposium on Interactive Ray Tracing.