

Performance Analysis of Machine Learning Classifiers for Fault Type Classification in Multi-Bus Power Systems

^{*1}Janak Raj Chadha, ²Megha Pandey, ³Tejash Sharma, ⁴N.R. Patne, ⁵Piyush Khadke

^{1,2,3,4}Department of Electrical Engineering, Visvesvaraya National Institute of Technology, Nagpur

⁵Shri Ramdeobaba College of Engineering and Management, Nagpur

Email: ¹janakrajchadha@gmail.com, ²meghapandey13@gmail.com, ³sharmatejash786@gmail.com,

⁴nrpatne@eee.vnit.ac.in, ⁵piyushkhadke88@gmail.com

Received: 06th October 2019, Accepted: 20th November 2019, Published: 31st December 2019

Abstract

In the past decade, multiple attempts have been made to classify faults on transmission lines using mathematical transformations and machine learning algorithms. The available literature primarily contains analysis of simple single line systems. This paper summarises the performance analysis of machine learning classification models on more complex power system models. Two different power system models are taken into consideration, a standard 3 bus 3 generator 2 load software model and a hardware simulator manufactured by SAMAR Instruments, Italy. The analysis of the first system captures the effects of multi-bus multi- generator systems on the performance of classifiers. The second system focuses on the results of testing the models on a hardware simulator. Discrete wavelet transform has been utilised for feature extraction and the performance of different classifiers has been analysed.

Keywords

Power Systems, Fault Classification, Machine Learning, Wavelet Transform, Multi-bus System, Classifiers

Introduction

There is some literature available on methods for fault detection, classification and location of faults in power systems using machine learning algorithms. K. Chen et al. [1] have comprehensively reviewed the methods currently used for these problems. However, in most cases, to the authors' best knowledge, the systems under consideration consist of a single transmission line with a generator on one side and a load on another.

While it makes intuitive sense to consider a single line model to train a model for a single relay, the presence of other lines, buses, generators and loads in a practical system affect the electrical quantities of the line under consideration. The first part of this paper tries to study the effects of this increase in system complexity on the performance of the algorithms.

The second part of this paper analyses the performance on a hardware simulator. Data of real electrical faults created on the simulator is used as a test data set while the training data set is created using a software model replica of the hardware as it has limited ports to create faults.

To extract useful information from the electrical signals, wavelet transform has proven to be a powerful transformation as it is able to capture both time and frequency domain information. In our analysis, we have considered features like the energy and the maximum coefficient of Discrete Wavelet Transform (DWT).

The principles of DWT and Machine Learning have been briefly discussed in section 2. The power system models have been described in section 3. Section 4 covers data set creation, feature selection & scaling. In section 5, the results have been summarised followed by the observations in section 6. The paper ends with a conclusion of work and future scope.

Discrete Wavelet Transform and Machine Learning Classification Algorithms

One might consider using aggregates (average, rms) of the electrical quantities as features. However, these quantities are not able to capture anything other than the magnitude of the signals. This may be problematic, especially in the case of symmetrical faults as the magnitude of a no-fault condition with a certain load may be close to the symmetrical fault threshold and will result in a false positive. Such problems are mitigated by applying mathematical transformations on the signals like the Discrete Wavelet Transform. Various classification algorithms can be used for fault classification, each with its own advantage. The performance of Support Vector Machines, k-Nearest Neighbours and Decision Trees have been compared.

A. Wavelet Transform and DWT

The wavelet transform is a mathematical transform that decomposes a signal into a variety of frequency bands and shows them as a function of time. The wavelet transform is useful for the fault signals which may not be purely sinusoidal or non-periodic[5]. Discrete Wavelet Transform (DWT) is the discrete version of wavelet transform and has a low-pass and a high-pass filter in the form of mathematical functions that are called scaling function and wavelet function respectively[6].

$$\text{Scaling function : } \Phi(x) = \sum_n h(n) \sqrt{2} \phi(2x - n) \quad (1)$$

$$\text{Wavelet function : } \Psi(x) = \sum_n h(n) \sqrt{2} \phi(2x - n) \quad (2)$$

In DWT the variety of frequency bands for the signal is achieved by passing the time-domain signal into a high-pass and low-pass filter iteratively. In the first iteration (level-1), the original time-domain signal is given to both the scaling $h_{\phi}[n]$ and wavelet function $h_{\psi}[n]$. In this process, half of the samples are eliminated according to Nyquist's criterion. This is because the highest frequency present after filtering is half of the highest frequency in the original signal. This whole process of filtering and sub-sampling can be iterated to get multiple levels. We get the coarse approximate and detailed coefficients from the low-pass and the high-pass filter respectively. These level-1 detailed coefficients are being used for classifying the faults.

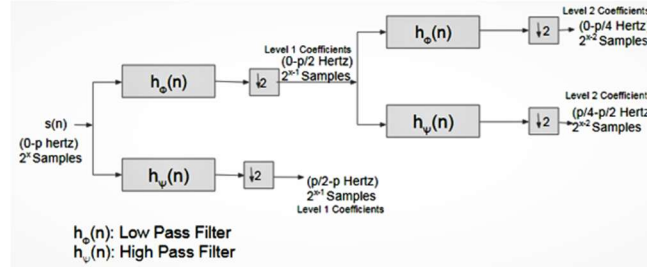


Fig. 1: A Two Level Filter Bank Block Diagram for DWT

B. Machine Learning Classification Algorithms

Classification algorithms are essentially a set of algorithms which are able to distinguish between different categories or groups based on the properties or features which make the groups easily distinguishable.

- 1) Support Vector Machines: It is a prominent machine learning model due to its ability to efficiently learn both linear and non linear classifications. An SVM finds an optimal boundary or hyper-plane (in an n-dimensional space) which separates the classes. The choice of hyper-plane is made such that its distance from the nearest data point on each side is maximised.
- 2) K Nearest Neighbours: A new instance is classified on the basis of the classes of its 'k' closest neighbours and the majority class is assigned. An odd 'k' is often a good choice to avoid ties in the majority vote. The closest neighbours can be selected based on different distance measures like the Euclidean, Hamming, or the Manhattan distance.
- 3) Decision Trees: Decision trees use a tree like structure to go from the independent variables to the targets in a recursively splitting fashion. The interior branch nodes represent the variables whereas the leaf nodes contain the categories. Various metrics (gini index/information gain etc.) are utilised for deciding the best root node variable and the order of splitting variables.

Systems Under Analysis

A. Three bus system

The three bus system consists of three star grounded generators with short circuit capacity of 100 MVA, 25 MVA and 50 MVA at bus 1, 2, and 3 respectively. Phase-phase voltage is 69 KV. Each of the generators has an X/R ratio of 7. Operating frequency is 50 Hz. Three phase RLC load connected to bus has an active load of 500 MW and reactive load of 10 MVar. Transmission line connected between the bus 1 and 3 (60 km) has been analysed. Fault data comprises of three phase voltages, three phase currents, and neutral current and is recorded at bus 1. The parameters of the line are given in table 1.

Resistance	0.11 ohms/km
Inductance	1.4 mH/km
Capacitance	9.45 nF/km

Table 1: Parameters of Transmission Line for Three Bus System

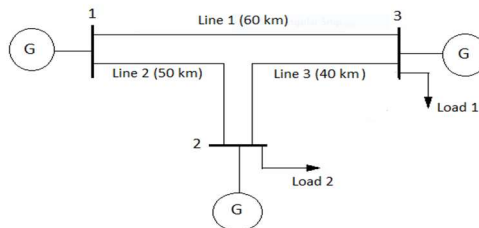


Fig. 2: Single Line Diagram of Three Bus System[8]

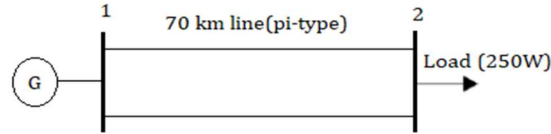
B. Hardware simulator system

The hardware simulator simulates two overhead Pi-model transmission lines (70 km each), at a voltage level of 120 kV or 220 kV with a variable load. The parameters of the line are given in table 2.

Resistance	0.229 ohms/km
Inductance	1.2 mH/km
Capacitance	9.43 nF/km

Table 2: Parameters of Transmission Line for Hardware Simulator

The two-bus power system model of the hardware simulator is also simulated using MATLAB for data set creation as explained in detail in section 4.2.2. Single line diagram of hardware simulator system is shown in figure 3.

**Fig. 3: Single Line Diagram of Hardware Simulator**

Feature Selection & Data Set Creation

A. Feature selection

Feature selection is extremely important for machine learning models. The result of applying DWT is a set of coefficients of size $N2^n$ for a discrete input signal where N is the total number of samples of the signal and n is the DWT level. Assuming a high sampling frequency for faithful reconstruction, the size of the set of coefficients of the relevant level will be at least a few hundred. While the entire set can be used as features, a large number of the coefficients are extremely low and provide little information. To overcome this problem, we have considered two metrics: 1. Energy of the set of coefficients, 2. Maximum coefficient of the set, where the Energy is defined as

$$E(x(i)) = \sum_{i=1}^{N2^n} x(i)^2 \quad (3)$$

The DWT was applied to the three phase voltages, three phase currents & neutral current. The energy and maximum coefficient for each of these signals were computed. Therefore, the total number of features were reduced from a few thousand ($7 * \text{a few hundred}$) to 7 as either energy or maximum coefficients. The results of the two sets of features have been compared.

B. Data set creation

1) Three bus system: Two training sets were simulated for the three bus system. In the first, only short circuit faults at different locations of the transmission line were considered. The second covers changes in fault resistances and different fault times (\equiv fault inception angles), both chosen arbitrarily. The values of the two sets are covered in tables 3 and 4.

Length of the line	60 km
Fault locations	Every 300m i.e. 300m,600m...60km
Types of fault	3 L-G, 3 L-L, 3 L-L-G, 1 L-L-L-G = 10
Training set size	200 locations x 10 types of faults = 2000

Table 3: Short Circuit Faults, Three Bus System

Length of the line	60 km
Fault locations	Every 500m i.e. 500m,1km...60km
Types of fault	3 L-G, 3 L-L, 3 L-L-G, 1 L-L-L-G = 10
Fault resistances	0, 5, 25 Ω
Fault inception angles	72° & 165°
Training set size	120 locations x 3 resistances x 2 inception angles x 10 types of faults = 7200

Table 4: Faults with Varying Resistances and Inception Angles

A test set was also simulated to test the performance of the classifiers trained on faults. To test if the classifiers generalized well, the test set needs to include situations not covered by the training set. To simulate real life situations as closely as possible, the fault locations, resistances and inception angles were selected randomly from a uniform distribution within a given range. A total of 50 combinations of randomly selected values of the three parameters were used and the ten types of fault were simulated for each combination, creating a total data set of 500 samples. The classifiers with the highest training accuracy based on a five-fold cross validation score were used to make predictions on the test set.

2) Hardware simulator system: A training and a test set were created for the hardware simulator system. The training set was simulated using MATLAB and the details are given in table 5. The test set was created by creating faults on the hardware simulator. To extract samples of the waveforms, a power analyser was used. The test set faults were created at the end of the transmission line and the data samples were obtained. A total of 5000 samples for 10 cycles of the signal were taken.

Length of the line	70 km
Fault locations	Every 500m i.e. 500m,1km...70km
Types of fault	3 L-G, 3 L-L, 3 L-L-G, 1 L-L-L-G = 10
Training set size	140 locations x 10 types of faults = 1400

Table 5: Short Circuit Faults, Hardware System

Results

A. Three bus system

The performance of different models has been compared based on the three aforementioned algorithms. For each algorithm, the results for a range of parameters and variants have been compared for a comprehensive study. For SVMs, the box constraint level i.e. C parameter and kernel scales i.e. γ parameter, have been covered. In the case of decision trees, the criterion for splitting and the maximum number of splits have been varied. For k nearest neighbours, the number of neighbours and the distance metric are varied. Multi-class classifiers have been created for each algorithm first, and the accuracy are calculated based on the confusion matrices obtained from the 5-fold cross validation process. The results are summarised in table 6.

Algorithm	Parameters with best performance	Cross validation score (5 fold)
SVM (One vs All)	Gaussian kernel: C = 1000, $\gamma=0.66$	91.7%
KNN (Weighted)	Distance metric: Cosine, k (neighbours): 10	91%
Decision Tree	Split criterion: Gini's index, max. Splits = 200	89.9%

Table 6: Accuracy Comparison of Multi-class Classifiers

Algorithm	Multi-class classifier score	Binary classifier score			
		R	Y	B	N
SVM (One vs All)	91.7	96.3	96.7	96.7	96.4
KNN (Weighted)	91	96.9	97.1	96.8	94.6
Decision Tree	89.9	94.8	96.1	95.7	95.5

Table 7: Multi-class vs Binary Classifiers Accuracy

For the top three performing algorithms in this case, the performance of the multi-class classifier has been compared with binary classifiers for each phase and neutral. The results are presented in table 7.

The top two binary classifiers, which are SVM and KNN, are used to make predictions. While it can be clearly seen that binary classifiers for each phase perform better than multi-class classifiers, a true check of performance is the comparison of the accuracy of the true class predictions. The individual classifiers' accuracy don't directly represent the results which are used for the protection system. Thus, the predictions from the binary classifiers are combined to obtain the true fault. Table

8 shows sample predictions of multi-class and binary classifiers and table 9 shows the prediction scores.

True class	Multi-class classifier prediction	Binary classifier prediction R Y B N	Combined binary classifier prediction
7(R-Y-G)	7	1 1 0 1	7
4(R-Y)	4	1 1 0 1	7

TABLE 8: Sample Prediction

Algorithm	Multi-class prediction score	Combined binary classifiers prediction score
SVM	88.4	85
KNN	88.6	87.8

Table 9: Test Set Prediction Scores

From these results, it can be seen that the multi-class classifiers have generalised well and there is a small difference between the training and test set scores. On the other hand, the test set accuracy of the combined binary classifiers are lower than the training set. This is due to the possibility of mis-classification after combining the results of the individual binary classifiers as shown in table 8.

B. Hardware simulator system

The performance of different sets of features and algorithms has been compared based on their cross-validation scores for the hardware simulator. The scores are summarised in table 10. The two sets of features compared are the energy of the coefficients and the maximum coefficient. In the considered algorithms, linear SVM was used with box constraint level (C) equal to 1000. For KNN, the number of neighbours (k) considered are 10 with cosine distance metric. For decision tree, maximum number of splits considered are 100.

Algorithm	Cross validation score (5 fold)	
	Energy	Maximum coefficient
SVM (One vs All)	99.8%	99.1%
KNN (Weighted)	67.6%	90.6%
Decision Tree	99.9%	98.9%

Table 10: Accuracy Comparison of Multi-class Classifiers

The models were tested on 10 different types of fault data obtained from the hardware simulator. The best results were obtained for decision tree. The results are presented in table 11.

Algorithm	Correct predictions	
	Energy	Maximum coefficient
Decision Tree	5/10	7/10

Table 11: Test Set Prediction Scores for Hardware System

Observations

A key observation from our analysis is the drop in training accuracy of algorithms on complex systems with multiple transmission lines connected to each bus. In particular, the drop in the three bus system is due to the mis-classification of the faults near the line ends i.e. close to the bus. Thus, the effects of the other lines connected at the bus and the generators/loads connected to those lines need to be taken into consideration while designing the training sets for complex power systems.

Another key observation is the misleading nature of individual classifiers for each phase and neutral. While their individual performance is better than multi-class classifiers, the true prediction results of the multi-class classifiers on the test set are better than those obtained by combining the predictions of each individual binary classifier.

For the software model of hardware simulator it was observed that cross validation score for energy was higher than the maximum coefficient in the two cases with comparable scores for each set of features i.e. SVM and decision trees.

The 3/10 mis-classified hardware simulator test faults were line-line faults. On inspecting these mis-classifications, it was found that this was due to a considerable difference in the neutral current values between the software model used for training

and the actual hardware model. Thus, the model mis-classified the line-line faults as line-line-ground faults.

As a system can be prone to unforeseen situations, one can look into approaches which can easily update the models based on the new data entries. One related approach, 'Online passive-aggressive algorithms' has been proposed by Crammer [9].

Conclusion

Based on our analysis and observations, we present certain conclusions and some suggestions for future work. From the analysis of the three bus system, it can be said with confidence that multi-class classifiers perform better than a combination of individual binary classifiers. The effects of other transmission lines and components (sources, loads) in the system cannot be ignored (especially near buses) while creating models and need to be analysed properly.

Through the analysis of the hardware simulator, it has been observed that there may be differences in the values in a software simulation and a real system. Values in a real system are often influenced by factors (environmental, weather etc.) which are mostly ignored in a simulation. Thus, software models either need to incorporate certain real life situations or work with data from real time simulators to create truly robust systems ready to work in parallel with current protection schemes.

Based on the observations, we conclude that fault type classification models for complex power systems require:

- More and randomised training data (not restricted to ranges of parameters)
- Analysis of effects of other lines and components in the system on the line under consideration
- Comprehensive test sets to ensure better generalisation of classifier models
- Testing with data from hardware simulators and real time simulators

References

- [1] K. Chen, C. Huang and J. He, "Fault detection, classification and location for transmission lines and distribution systems: a review on the methods," in *High Voltage*, vol. 1, no. 1, pp. 25-33, 4 2016.
- [2] A. Ngaopitakkul, C. Pothisarn, S. Bunjongjit and B. Suechoey, "An Application of Discrete Wavelet Transform and Support Vector Machines Algorithm for Classification of Fault Types on Underground Cable," 2012 Third International Conference on Innovations in Bio-Inspired Computing and Applications, Kaohsiung, 2012, pp. 85-88.
- [3] M. Singh, B. K. Panigrahi and R. P. Maheshwari, "Transmission line fault detection and classification," 2011 International Conference on Emerging Trends in Electrical and Computer Technology, Tamil Nadu, 2011, pp. 15-22.
- [4] S. R. Samantaray, "Decision tree-based fault zone identification and fault classification in flexible AC transmissions-based transmission line," in *IET Generation, Transmission & Distribution*, vol. 3, no. 5, pp. 425-436, May 2009.
- [5] Abhisek Ukil, Rastko Ivanovi, "Abrupt change detection in power system fault analysis using adaptive whitening filter and wavelet transform," in *Electric Power Systems Research*, vol. 76, issues 9-10, pp. 815-823, 2006.
- [6] Zhiguo Chen, "A new high-speed foreign Fibre detection system with Machine vision," in *Mathematical Problems in Engineering*, 2010.
- [7] F. A. Albasri, A. R. Alroomi and J. H. Talaq, "Optimal coordination of directional overcurrent Relays Using Biogeography-based optimization algorithms," *IEEE Transactions on Power Delivery*, 2015, 30, pp. 1810-1820.
- [8] Koby Crammer, Ofer Dekel, Joseph Keshet, Shai Shalev-Shwartz, and Yoram Singer, "Online passive-aggressive algorithms," in *Journal of Machine Learning Research*, 2006, pp. 551-585.
- [9] K. M. Silva, B. A. Souza and N. S. D. Brito, "Fault detection and classification in transmission lines based on wavelet transform and ANN," *IEEE Transactions on Power Delivery*, 2006, Vol. 21, no. 4, pp. 2058-2063.
- [10] S. M. Shahrtash and A. Jamehbozorg, "A decision tree based method for fault classification in transmission lines," 2008 IEEE/PES Transmission and Distribution Conference and Exposition, Chicago, IL, 2008, pp. 1-5.
- [11] V. Malathi and N. S. Marimuthu, "Multi-class Support Vector Machine approach for fault classification in power transmission line," 2008 IEEE International Conference on Sustainable Energy Technologies, Singapore, 2008, pp. 67-71.
- [12] C. H. Kim and R. Aggarwal, "Wavelet transform in power system," in *IET Power Engineering Journal*, Vol. 15, August 2001, pp. 193-200.
- [13] M. M. Mansour, S. F. Mekhamer and N. El-Kharbawe, "A Modified Particle Swarm Optimizer for the Coordination of Directional Overcurrent Relays," *IEEE Transactions on Power Delivery*, 2007, 22, pp. 1400-1410.