Random Forest Regressor (RFR) based approach for Detecting Fault Location and Duration Anomalies in Power Systems

Zakaria El Mrabet[1], Niroop Sugunaraj[1], Prakash Ranganathan[1], Shrirang Abhyankar[2]
[1]University of North Dakota, Grand Forks, ND-58202
[2]Electricity Infrastructure and Buildings Division,
Pacific Northwest National Laboratory, Richland, WA 99354

PNNL

C2SR
CENTER FOR CYBER SECURITY RESEARCH

DECS

NSF

Abstract

Power system failures or outages due to short-circuits or "faults" can result in long service interruptions leading to significant socio-economic consequences. It is critical for electrical utilities to quickly ascertain fault characteristics, including location, type, and duration, to reduce the service time of an outage. In this project, we propose a data-driven random forest regressor (RFR)-based model to detect real-time fault location and its duration simultaneously. Four cases were studied to evaluate the performance of RFR and. RFR consistently outperformed other models in detection accuracy, prediction error, and processing time.

Introduction

Fault identification is critical for seamless power grid operation. Utilities are working around the clock to reduce outage rates from interruptions such as contact with natural vegetation, animals, or weather events. The unplanned outages can lead to long service interruptions and significant economic impact to the customers. Machine learning (ML)-based approaches for detecting fault locations have been reported in peer-reviewed literature [1]. Compared with detecting fault location, relatively few works have been carried out to predict fault duration. Fault location detection is a multiclass classification problem (fault position) while fault duration prediction is regarded as a regression problem as the output would be a continuous value (fault duration).

Existing Works

Table 1. Existing literature on fault anomaly detection [2-4].

Category	Approach	Advantages	Limitations
Conventional	Impedance-based	Ease of Implementation	Fault duration not considered
	Time wave-based	Large system parameter values	Accuracy dependent on input values. Fault duration not considered
Machine Learning	NN-based	High tolerance to noise	Inappropriate for streaming environments
	RF	High accuracy (90.9%)	Fault duration not considered
	CNN-based	Optimal localization estimation	Fault duration not considered
	kNN	High accuracy and low error margins	Trained/tested on PV only
	HAT-based	High classification accuracy for single and multi-class cases	Fault location and duration not considered

Experimental Methodology

Dataset: The simulated fault scenarios were completed using GridPACK software, an open-source framework designed to support the development and implementation of power grid applications. Examples of these applications include power flow simulations for the electric grid, contingency analysis of the power grid, state estimation based on electric grid measurements, and the dynamic simulation of the power grid. These applications can run on high-performance computing (HPC) architecture. The total number of samples for all simulated scenarios equaled 53,512 samples (37,458 samples for training and 16,053 samples for testing).

Model: Random forest F is an ensemble approach with several independent and uncorrelated decision trees *F = [t1, t2, . . . , tt]*. These uncorrelated trees assist model F in achieving an accurate generalization by injecting randomness into the decision trees. These generalizations rely on the application of a bagging technique, which combines the concepts of bootstrapping and aggregation. Ensemble methods like random forest are effective in detecting concept drifts i.e., unexpected events that occur in a streaming PMU environment (see Figure 2 below).

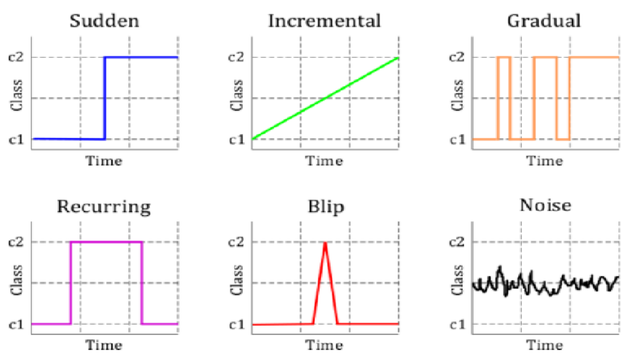

Figure 2. Types of concept drift.

Evaluation Metrics

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Mean Absolute Error (MAE)} = \left|\frac{1}{n}\sum_{k=0}^{n}(y' - y)\right|$$

$$\text{Mean Squared Error (MSE)} = \left|\frac{1}{n}\sum_{k=0}^{n}(y' - y)^2\right|$$

Data Driven Workflow

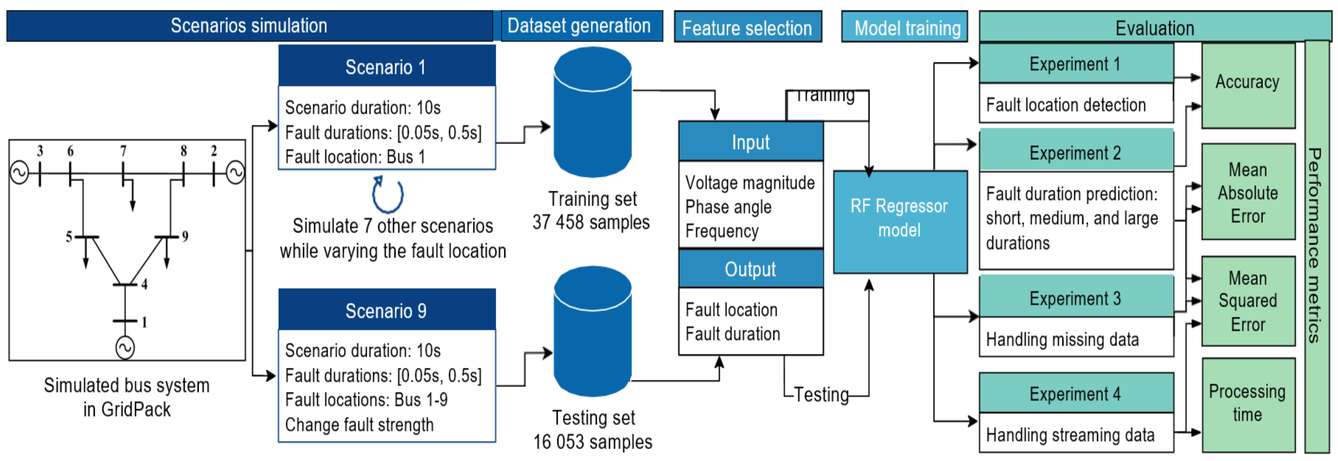

Figure 1. Input, processing, and evaluation for fault detection.

Experimental Results

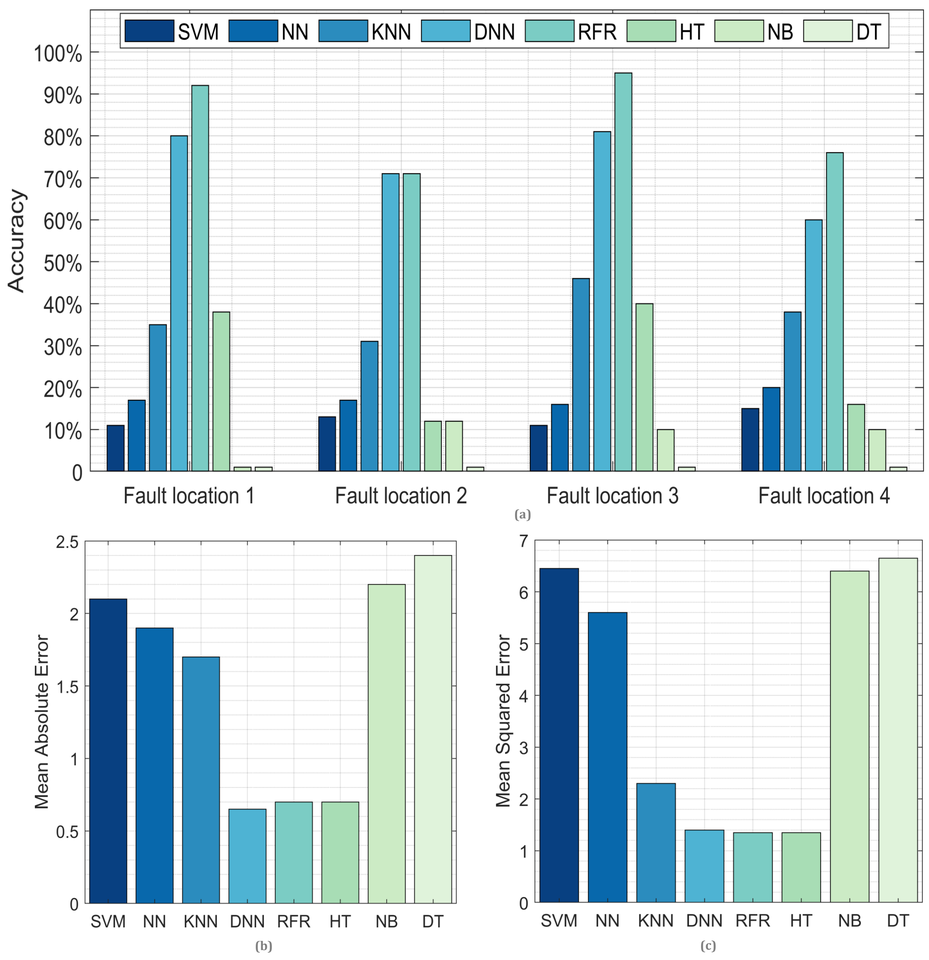

Figure 3. Performance evaluation for multiple models in terms of (a) accuracy (b) mean absolute error (MAE), and (c) mean squared error (MSE).

Table 2. Hyper-parameters for trained models.

Model	Hyper-parameters
RFR	Max features, number of trees
KNN	Weight function (uniform or distance)
SVM	Kernels (polynomial or radial basis function)
DT	Minimum leaf size, random state
DNN	Activiation function, no. of hidden nodes, no. of hidden layers
HT	Split function (Gini index, information gain) and split confidence
NN	Activation function (ReLU, identity) and no. of hidden nodes
NB	Alpha coefficient, lambda

Conclusions

A random forest regression (RFR)-based model was successfully implemented to identify the location and duration of faults. Various fault scenarios were modeled using PNNL's GridPACK software for the generation of the training dataset. A total of nine fault scenarios was simulated by injecting faults on specific buses over a specified period and tested to detecting fault location and predicting fault duration. A comparison was also conducted between the RFR model and several state-of-the-art models using multiple performance metrics, including accuracy, MSE, and processing time. Results indicate that both RFR and DNN models can detect the location and duration of a fault with an accuracy of 84% and 72%, respectively. The RFR, DNN, and HT models yielded better results when predicting faults in streaming networks. Overall, the RFR model consistently outperformed the other models, making it appropriate for real- time situational awareness deployments to determine both the location and duration of the faults while handling missing data.

Contact Information

Niroop Sugunaraj
Doctoral Student
University of North Dakota
niroop.sugunaraj@und,edu

Acknowledgements

Research presented in this poster was supported by the National Science Foundation under NSF (Award #1537565.)

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

The findings in this poster were published in MDPI Sensors. Please cite using: El Mrabet, Z., Sugunaraj, N., Ranganathan, P., & Abhyankar, S. (2022). Random Forest Regressor-Based Approach for Detecting Fault Location and Duration in Power Systems. Sensors, 22(2), 458.

References

1. *Joyokusumo, I.; Putra, H.; Fatchurrahman, R. A Machine Learning-Based Strategy For Predicting The Fault Recovery Duration Class In Electric Power Transmission System. In Proceedings of the 2020 International Conference on Technology and Policy in Energy and Electric Power (ICT-PEP), Bandung, Indonesia, 23–24 September 2020; pp. 252–257.*
2. *Gururajapathy, S.S.; Mokhlis, H.; Illias, H.A. Fault location and detection techniques in power distribution systems with distributed generation: A review. Renew. Sustain. Energy Rev.* ***2017****, 74, 949–958. [CrossRef]*
3. *Ma, G.; Jiang, L.; Zhou, K.; Xu, G. A Method of line fault location based on traveling wave theory. Int. J. Control Autom.* ***2016****, 9, 261–270. [CrossRef]*
4. *Zainab, A.; Refaat, S.S.; Syed, D.; Ghrayeb, A.; Abu-Rub, H. Faulted Line Identification and Localization in Power System using Machine Learning Techniques. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 2975–2981.*

UND
UNIVERSITY OF NORTH DAKOTA

UND School of Electrical Engineering & Computer Science
UNIVERSITY OF NORTH DAKOTA