Machine Learning Methods for Power Line Outage Identification

Jia He Maggie X. Cheng

Abstract

As Phasor Measurement Units (PMUs) become widely deployed, power systems can take advantage of the large amount of data provided by PMUs and leverage the advances in big data analytics to improve real-time monitoring and diagnosis. In this paper, we develop practical analytics that are not tightly coupled with the power flow analysis and state estimation, as these tasks require detailed and accurate information about the power system. We focus on power line outage identification, and use a machine learning framework to locate line outages. The same framework is used for the prediction of both single line and multiple line outages. We investigate a range of machine learning algorithms and feature extraction methods. The algorithms are designed to capture the essential dynamic characteristics of the power system when the topology change occurs abruptly. The proposed methods use only voltage phasor angles obtained by continuously monitoring the buses. We tested the proposed methods on their prediction performance under different levels of noise and missingness. It is shown that the proposed methods have better tolerance for noisy data and incomplete data when compared to the previous work that involves solving power flow equations or state estimation equations.

Index Terms

Power Systems, Line Outage, Machine Learning, Logistic Regression, Random Forest

I. Introduction

Following the major blackout in 2003, topics on anomaly detection and subsequent fault localization have been heavily investigated. Among all disastrous events, power line outage hits most frequently. Power line outage identification becomes the first and most powerful tactic to improve situational awareness. Being aware of the changes on the power lines is paramount for

This work was supported in part by the National Science Foundation of USA under grants 1936873, 1854077 and 1854078.

J. He and M. Cheng are with Illinois Institute of Technology, Chicago, IL 60616 USA (e-mail: maggie.cheng@iit.edu).

several critical tasks, such as state estimation, power load flow analysis, real-time contingency analysis [1]. For instance, the state matrix used in the state estimation is based on an assumed topology and parameters. A topological change of the grid can completely overthrow the state estimation [2]. Topology errors caused by power line outages can potentially lead to very costly large-scale power outages if not addressed timely [3], [4].

This work focuses on identifying the line outage location(s) and we consider it a classification problem in machine learning. We investigate several classification algorithms, including logistic regression, random forest, and graph convolutional neural networks. The methods can be applied in real-time. Although training a classifier can be time-consuming, the prediction of power line status can be done in real-time.

Previous work on line outage detection and identification heavily rely on other critical tasks of the power system, such as state estimation, or load-flow analysis [1]. Some either use the residual of state-estimation [2], or carry out a joint outage identification and state estimation [5]. However, these tasks require a lot of global information that may not be available all the time. Measurement data from PMU are more accessible. In this paper, we use measurement data only to infer the status of power lines. The advantage of this approach is that there is no interlocking with other tasks, and therefore the line outage identification can be used independently and prior to the implement of state estimation or power flow analysis.

The data-driven inference approach compared to the previous work that relies on accurate state information of the system for line outage detection/identification is a fundamental breakthrough. Compared to earlier works that use machine learning approaches, the proposed work distinguishes itself in several ways: 1) It is not based on a particular flow model (DC or AC). The input to the classifiers is derived information from PMU measurements, and the classifiers do not use model-specific information; 2) We do not use voltage magnitudes and power injections. We only use phasor angles. Instead of using phasor angles directly, we use the difference between the phasor angles from two adjacent buses to build time series. We then use the spectrum features extracted from the time series as input for the classifiers. These features more accurately capture the correlation with the network topology than those used in previous work; 3) The performance of the classifiers are reported by using precision and recall as metrics, instead of an overall accuracy or error rate as in the past work. The overall accuracy is high mainly because the classifiers perform well for the non-case, i.e., it predicts the network has no outage when there is no outage. For rare event detection, the non-case is the majority case in the dataset. Even

though a classifier performs poorly for the ”yes” case, the overall accuracy is still high. However, we think it is a reasonable argument that to detect the ”yes” case accurately is more important. We therefore emphasize how to improve the detection rate (recall) in this paper; 4) Our work does not depend on the sparsity assumption as in [1]. Since each line is predicted separately, the number of line outages that it can predict is unlimited; 5) Using our method, the classifiers do not require PMU data from all buses. Experiment results show that it requires only a subset of PMU data. The algorithms are robust to noise and missingness. 6) Finally, to our knowledge, this is the first work to predict power line outage in a very large system (1354-bus with 1991 lines) at high detection rates.

The rest of the paper is organized as follows: in Section II, we survey the most related work in power line outage detection and identification; in Section III, we cover the preliminaries underlying the proposed method, and describe the characteristics of the signal caused by line outage; in Section IV, we provide detailed description of the line identification methods; in Section V, we provide extensive simulation results on single and multiple line outages using the 39-bus, 118-bus IEEE standard test systems and a very large system; in Section VI, we conclude the paper with outlook for future work.

II. Related Work

As PMUs become increasingly deployed, power line outage detection and identification based on PMU measurements have received much attention recently. In early work [6], system topology information together with PMU phasor angle measurements are used to detect a single line outage. To further determine the broken line, the pre-outage flow is estimated, in which power flow equation is used and the system admittance matrix is needed.

Tate and Overbye further extended the single line outage detection method in [6] to double line outage detection [7]. It is proved in [7] that there exists indistinguishable outages due to incomplete PMU deployment, and a method of recognizing these indistinguishable outages is presented. This conclusion is also consistent with the conclusion of our own study. The indistinguishable outages due to the limited PMU deployment is a major contributing factor of prediction errors.

Other recent works on outage identification based on PMU data include [8] and [1]. In [8], a linear multinomial regression model is estimated by solving a maximum likelihood problem using the sampled PMU data, however, the classifier is only effective in the presence of perfect

dynamic simulation. [1] focuses on detecting multiple line outages at low complexity. [1] solved the sparse line outage identification problem by using the DC linear power flow model. Like [1], the proposed logistic regression and random forest models also avoid the combinatorial complexity issue; but different from [1], the proposed method only uses PMU data and does not need the reactance information on the power lines.

There are also notable works that solve outage detection, state estimation and optimal sensor locations as coupled problems. In [5], a joint detection and estimation problem was studied for outage identification in power systems. The authors employed a Bayesian framework, in which the prior distributions of outage events and network states are assumed to be Gaussian, and developed the closed-form joint posterior distribution of outage events and network states, which can be used to determine the optimal joint outage detector and state estimator. It is also not rare to address line outage detection and identification as one coupled problem (see [9], [10]). Both [9] and [10] use incremental changes from measurements between two time instances as random variables to build time series. [9] uses phasor angles and [10] uses voltage magnitudes. In [9] change point detection is performed based on the time series directly, and line outage is detected from the first series that reported the change. In [10] detection and identification are based on changes on conditional correlation. There are also several works using message passing algorithms for network topology inference, e.g., [11], [12].

In recent years, there are a handful of works using machine learning approaches for power line outage detection. In [13], a classical support vector machine (SVM) approach is developed for single line outage detection, and it was proved successful on a 14-bus system. In [14] and [15] neural networks are developed to predict single and double line outages, and both use a single hidden layer. In [14], the input to the neural network includes phasor angles and power injections; in [15] the input is voltage magnitudes and phasor angles. [15] addressed partial data inference and [16] discussed optimal locations for PMUs. It was solved as an optimization problem to maximize the system's ability to detect single-line outages.

III. Line Outage Signal

This paper focuses on the identification of power line outage, i.e., to locate which lines are out. We use the measurement data from buses for this task. Such data are available from Phasor Measurement Units, which are widely available nowadays. Throughout the implementation of

the method, we assume no prior knowledge about the admittance matrix or state equations of the system.

Early work has proposed a method of change point detection from time series [17], which can be used to detect changes from time series of measurements, but does not tell exactly which lines are out. To further determine whether it is single line outage or multiple line outages, and identify which lines are out, we formulate the problem as a classification problem, and use different machine learning algorithms to solve the problem.

The first question to answer is: among the rich data collected from PMUs, which ones are useful for classification? It is observed that in the normal operation when two buses are connected by a transmission line, the power transfer amount is approximately proportional to the angle difference between the two buses. When the line between two buses is open, the angles at two buses lose this dependent relationship. Therefore upon line breakage, we expect to see a transition from a relatively stable signal to a more dramatically changing signal when we measure the angle differences. The machine learning algorithms can learn this feature and identify which line is out.

Fig. 1 shows the angle difference between bus 1 and bus 39 when line outages happen in the IEEE 39-bus system. When a line outage occurs at a branch, we observe a change in the difference of angles between the two buses that the branch connects, as well as from other pairs of buses connected by a branch. Although the change in the magnitude is small, the change in the dynamic feature is significant — a high frequency component occurred at the time of the abrupt change. Fig. 2 shows the frequency domain features. Although the plots look similar in shape, the differences in amplitudes across rows are more significant than the differences across columns, which suggests that the amplitudes of the dominant frequencies under different line outage scenarios are very different, more so than the differences caused by varying loads under the same topology.

Based on the observation above, the only information needed for the purpose of classification is the phasor angles at buses, and more specifically, it is the angle difference between a pair of connected buses that varies with the line outage location. Assume there are angle measurements from N buses, then there is a N-dimensional time series $\{\boldsymbol{\theta_t}\}$, $t = 1 \dots T$, where $\boldsymbol{\theta_t} \in \mathcal{R}^N$. We will not directly use $\{\theta_i\}$ as input data, instead we define $\theta'_i = \sum_{(i,j)\in\varepsilon} \theta_i - \theta_j, \forall i$ where ε is the set of lines in the power grid network, and uses $\{\theta'_i\}$ as input data. We could use $\{\theta_i - \theta_j\}, \forall (i,j) \in \varepsilon$

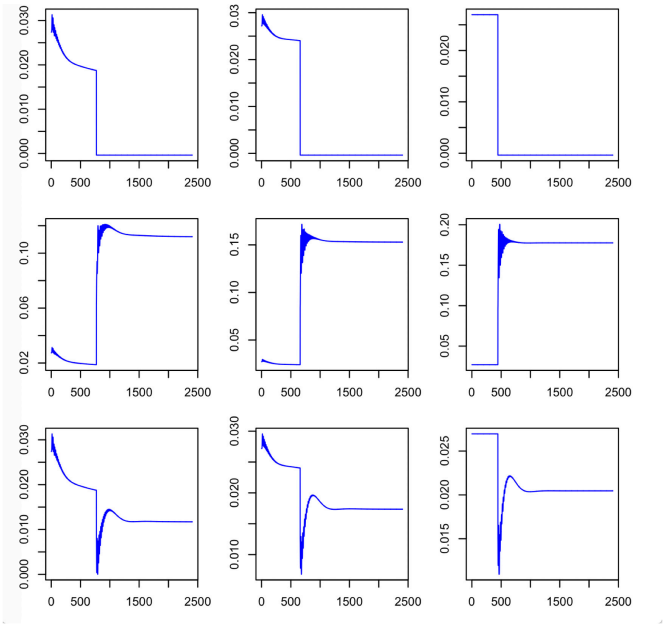


Fig. 1: Observed time series for $\theta_1 - \theta_{39}$ when a line outage happens. X-axis is time, Y-axis is the angle difference in radians. Top row: when line (1, 2) is out; middle row: when line (1, 39) is out, bottom row: when line (2, 30) is out. The three columns correspond to three different load settings.

as input, however, this leads to very high dimensional feature space since the number of lines is much higher than the number of buses in a graph. Using the sum of the angle differences is sufficient to capture the signal pattern with a minimum dimension in feature space.

IV. Line Outage Identification

We consider using three types of classification algorithms for the task of line outage identification.

- Logistic Regression (LR)
- Random Forest (RF)
- Graph Convolutional Network (GCN)

A. *Logistic Regression*

Logistic regression ([18]) is used to model binary response variable, Y. Given an observation X, we can compute the probability of Y being one. If $\mathbb{P}(Y = 1) > 0.5$, we predict the line is out.

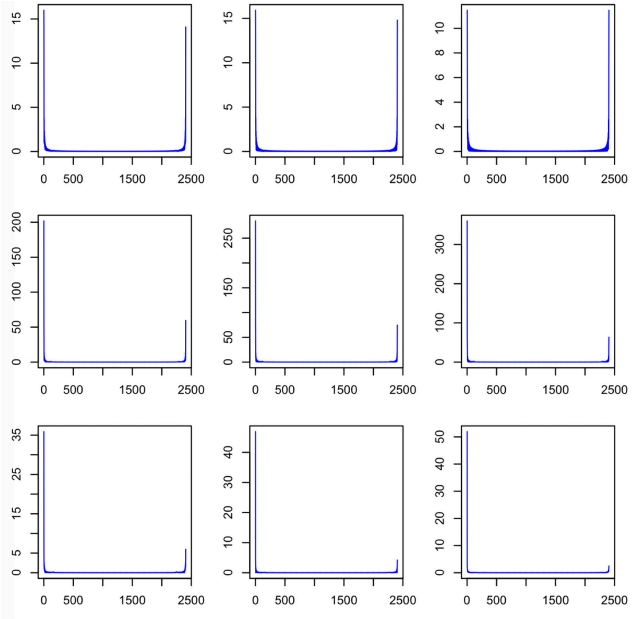


Fig. 2: The spectrum of the angle difference in Figure 1. X-axis is the frequency index, and Y-axis is the amplitude.

The model is trained by using pairs of (Y, X). The detail of the logistic model is included in the Appendix A. A separate logistic model is estimated for each line. If there are L lines, there are L models. For line outage detection, we have a vector with the l-th element $p^{(l)} = \mathbb{P}(Y^{(l)} = 1)$ indicating the probability of line l being open.

The logistic regression method is used for both single line outage and multiple line outages. The process involves training each model and then predicting $p^{(l)}$ for the new data, for $l = 1, \ldots, L$. The computational complexity does not increase with the number of outages, thus making the algorithm scale better than the multiclass classification methods.

B. Random Forest

Random Forest is a tree-based method [19]. Different from regression models, Random Forest uses decision trees as building blocks to construct prediction models.

A classification tree here is used to predict a qualitative response (Yes or No). Using decision trees for classification essentially involves two steps: In the first step — training, we divide the predictor space defined by the set of all possible values for the m-dimensional predictor variables

into J distinct and non-overlapping high-dimensional regions $\{R_1, \ldots, R_J\}$, then in the second step— prediction, for a new observation that falls into the region R_j, we assign it to the most commonly occurring class of training observations in R_j. In the training step, the regions are constructed to minimize the classification error rate, i.e., the fraction of the training observations in that region that do not belong to the most common class. In practice, either cross-entropy or Gini Index is used to evaluate the classification error rate.

Random forest involves producing multiple trees, which are then combined to yield a single consensus prediction by averaging all the predictions. When building individual decision trees, each time a split of the predictor space is considered, a random sample of p predictors are randomly chosen as split candidates out of the full set of m predictors. The split then uses only one of the p predictors. p is typically an integer smaller than $\sqrt{m}$.

Using random forest for the classification of line outage can be carried out in two ways:

1) Binary outcome. For each possible line outage, we train a separate model to decide if the line is broken, and then apply each of the models to the new observation. If multiple models yield "Yes", then there are multiple line outages. The complexity of the problem does not grow with the number of broken lines.
2) Multi-categorical outcome. We build one training model for all possible classes. If there are L possible line outages, there are $L + 1$ classes, or $L + 1$ possible outcomes, with outcome$= 0$ indicating no line is broken, outcome$= l$ indicating line l is broken. This method is only suitable for single line outage, as in the case of multiple line outages, the number of classes grows exponentially with the number of broken lines.

Experiments on single line outage detection indicate that multi-categorical random forest underperforms binary random forest by at least 20% in detection rate. For multiple line outage detection, the multi-categorical version cannot even detect 50% of the line outages due to having too many categories. Having too many categories with limited training data is detrimental to the method. Therefore, in this paper, we use random forest with binary outcomes, which yields not only less complexity but also better performance.

C. Graph Convolutional Network

The graph convolutional neural network has two essential components that distinguish itself from other convolutional neural networks: Graph Fourier Transform (GFT) of the signal, and

graph convolution operation in the neural network. The GFT step captures the spatial feature, and the convolution step captures the temporal feature.

Let matrix $\mathbf{GFT} = \mathbf{V}^{-1}$, where $\mathbf{V}$ is the eigenvector matrix of the standardized adjacency matrix $\tilde{A}$ (see Appendix B). Then the graph Fourier transformation of a vector $\boldsymbol{s}$ is given as a matrix-vector multiplication:

$$\widehat{\boldsymbol{s}} = \mathbf{GFT}\boldsymbol{s} \tag{1}$$

The output of Graph Fourier Transform provides some notion of frequency on graphs. In discrete Fourier transform, the notion of frequency is straightforward: frequency $f_k = \frac{2\pi k}{N}$, which is interpreted as k cycles per N samples, and large frequency corresponds to highly variated spectral components. Similarly, we can define high graph frequencies as those whose spectral components have a high variation, i.e., magnitudes of neighboring nodes are highly variated. It is proved in [20] that for adjacency matrix, smaller eigenvalues corresponds to higher variated eigenvectors, thus represents higher frequencies.

1) Spectral Convolution in GCN: For our application, the graph under consideration is undirected. Since the normalized adjacency matrix $\tilde{A}$ is symmetric, we have $\mathbf{GFT} = \mathbf{V}^{-1} = \mathbf{V}^T \in \mathcal{R}^{N\times N}$.

To derive output features from the convolution layer, we first define the frequency filter as a degree H polynomial in matrix $\mathbf{\Lambda}$ to consider information from nodes within H hops.

$$w(\mathbf{\Lambda}) = \sum_{h=0}^{H} w_h \mathbf{\Lambda}^h \tag{2}$$

The convolutional output is then given by

$$\widehat{\boldsymbol{y}} = w(\mathbf{\Lambda})\widehat{\boldsymbol{x}} = w_0\widehat{\boldsymbol{x}} + w_1\mathbf{\Lambda}\widehat{\boldsymbol{x}} + \ldots + w_H\mathbf{\Lambda}^H\widehat{\boldsymbol{x}}, \tag{3}$$

where $w_i, i = 0, 1, \ldots, H$, are trainable weights, and $\widehat{\boldsymbol{x}} = \mathbf{V}^T\boldsymbol{x}$. Elements of $\widehat{\boldsymbol{x}}$ are the spectral coefficients of $\boldsymbol{x}$. In spectral convolution, we take $\widehat{\boldsymbol{x}}$ as input features for the convolution layer.

The generalization of the above operation to a middle layer in the GCN, where there are G output features and K input features, is achieved by using a simple summation,

$$\widehat{\boldsymbol{y}}^g = \sum_{k=1}^{K} w^{k,g}(\mathbf{\Lambda})\widehat{\boldsymbol{x}}^k \tag{4}$$

where each parameter $w^{k,g}(\mathbf{\Lambda})$ is used to connect the kth input feature $\widehat{\boldsymbol{x}}^k$ to the gth output feature $\widehat{\boldsymbol{y}}^g$. Same as convolution in spatial domain, the number of learnable weights is $\mathcal{O}(KG)$.

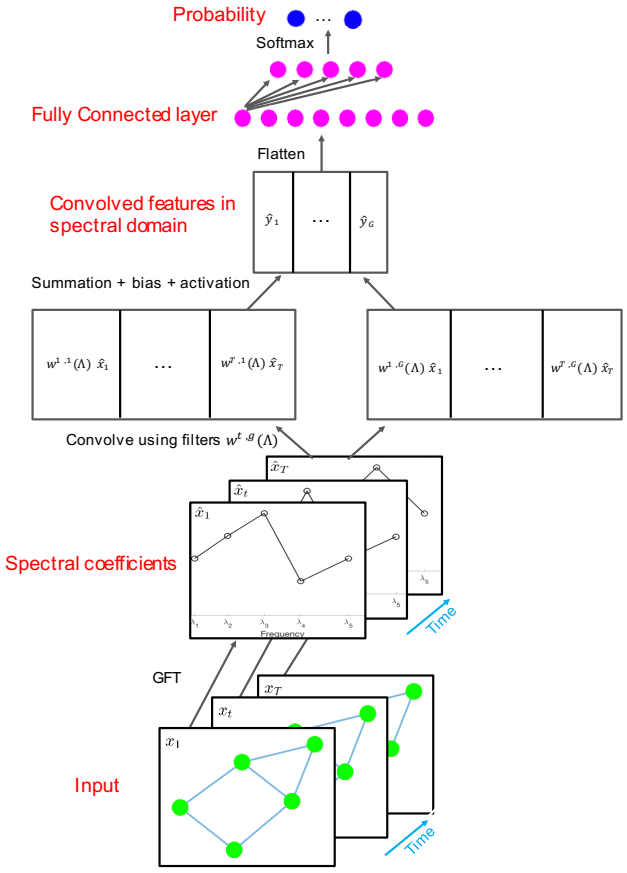


Fig. 3: Graph Convolutional Network model architecture. GFT is first applied to $\boldsymbol{x_t}$ to obtain graph frequency representation $\widehat{\boldsymbol{x}}_t$ of the signals. Then convolution in frequency domain is carried out by defining filters as a linear combination of graph frequencies matrix $\boldsymbol{\Lambda}$.

The input to the GCN is a graph signal $\{\boldsymbol{x_t}, t = 1, \ldots, T\}$, with each $\boldsymbol{x_t} \in \mathcal{R}^N$. The GFT operation is first applied to each $\boldsymbol{x_t}$ to get $\widehat{\boldsymbol{x}}_t$, and then the convolved features are obtained using (4). To complete the convolution layer, a point-wise nonlinear activation function is applied to each component of $\widehat{\boldsymbol{y}}$. The corresponding architecture is shown in Figure 3.

D. Feature Extraction

The data matrix $\{\boldsymbol{\theta}'\}$ is a matrix of $N \times T$. $\{\boldsymbol{\theta}'\}$ cannot be directly used as input for classification yet. The elements of the matrix are low level features for classification. Such low level features won't make a good predictor, since each element is a local feature, and it

doesn't capture the spatial feature across the network. The successful identification of line outage needs to recognize the network-wide features.

For each of the three proposed methods, we have different ways to overcome the issue of local features. For GCN, the Graph Fourier Transform operation serves this purpose. As shown in Fig. 3, the output of GTF is used as input for the convolutional layer. For LR and RF, we extract the network-wide dynamic features from the frequency domain by using Fourier Transform, and the output of FFT is used as input for LR and RF algorithms.

Fast Fourier Transform (FFT) takes a discrete signal in the time domain and transforms that signal into its discrete frequency domain representation. Assume the input signal is a length-n vector $\{x_j\}, j = 0, \ldots, n-1$. The output of FFT is

$$X_k = \sum_{j=0}^{n-1} x_j e^{-i2\pi kj/n}$$

where X_k is a complex number that encodes both amplitude and phase of a complex sinusoidal component $e^{i2\pi kj}$ of function x_j, and the frequency of the component is k cycles per n samples.

FFT results in a sequence of $\{X_k\}$, with $k = 0, \ldots, n-1$. We consider the dominant non-zero frequencies (i.e., the frequencies with the largest amplitudes) so that the dynamic features of the time series are fully captured in the frequency domain. In this paper, we use the amplitudes $|X_k|$ corresponding to the dominant frequencies $\{k\}$ as predictor variables for the classification problem.

The dominant frequencies can be selected by using the critical value from a Gumbel distribution (see Appendix C). A simpler approach is to sort the amplitudes and then select the top K (e.g., $K = 3$ worked well in this application) dominant non-zero frequencies. Although a fixed value K may correspond to a different significance level α, when it comes to line outage identification, the results from the two approaches are very close and non-distinguishable.

The extracted spectral feature compared to angle time series have the following advantages: 1) has lower dimension, therefore it requires less training time; 2) more accurately captures the relative changes of the angles and is independent of the angle measurements, therefore it is more stable when the load changes.

V. Results

A. Experiment Setup

The machine learning models are tested on IEEE standard test systems. Time domain simulation of the systems is implemented in Power System Analysis Toolbox (PSAT) [21]. PSAT is a free open source package equipped with modules for solving power flow (PF), optimal power flow (OPF), continuation power flow (CPF) and time domain simulation (TDS). In this paper, we do not use power flow analysis, since we assume neither the admittance matrix nor the Jacobian matrix is available. We use the time domain analysis part of PSAT, and the only information needed for the proposed methods is the incidence matrix for the line topology and the PMU measurements from buses. The OMP method from [1] will need the line parameters (e.g., the reactance along the line) to work.

We simulate the power system in PSAT and collect phasor angles $\{\theta_i\}$ at each bus. Each simulation lasts for a total of 200s, and the line outage occurs at 100s. We collect the data at the interval of 1/8 second or at a frequency of 8 samples per second. Single line outage is simulated in IEEE 39-bus and 118-bus systems, and multiple line outage is simulated in the 1354-bus system. The reason for using a larger system for multiple line outage detection is that we need to ensure the network is still connected after the line outages. If the network is disconnected with an isolated island, some buses will be disconnected from the rest of the network, and bus voltages will have no variation after the outage, thus the problem of identifying broken lines becomes too trivial.

The dataset for the 39-bus system has a total of 460 data points, among which 100 cases are line outage cases, and others are non-outage cases. The dataset for the 118-bus system has a total of 1120 cases with 100 line outage cases and the rest is non-outage cases. Each case is generated under a random initial bus load setting. The dataset is randomly split into a training set and a testing set following a 80%-20% rule.

The PSAT simulation provides measurements at buses as time series. We first use PMU measurements from all buses for training and testing, then we test on the situations where not all buses have PMUs installed, and therefore we have to deal with missing data. In addition, random Gaussian noise is added, and the robustness of the algorithms against noise is tested.

For GCN, the architecture consists of a convolution layer, followed by a readout layer, a fully-connected layer and a softmax activation function for classification. The convolution layer

uses up to one-hop neighborhood information. In this application, it is not necessary to use information from two or more hops although the GCN model is flexible to consider a larger neighborhood. The Adam optimizer is used for optimization, and exponential decay is applied to the learning rate, with a starting rate 0.001 and a decay rate 0.9. Cross entropy is used for the loss function. The number of epochs during training is 50 with 60 examples per batch.

B. Performance Metrics

We use *recall* and *precision* to assess the performance of the machine learning algorithms. Precision is defined as the fraction of true outages that are detected among the total detected outages, whereas recall (i.e., detection rate) is the fraction of true outages that are detected among all true outages.

$$\text{precision} = \text{true positives}/(\text{true positives} + \text{false positives})$$

$$\text{recall} = \text{true positives}/(\text{true positives} + \text{false negatives})$$

For rare event detection, an algorithm with a low detection rate can still achieve a high overall accuracy if the negative cases are dominant in the dataset and the algorithm produces a perfect true negative rate. Power line outage is a rare event, therefore recall is a better performance indicator than the overall accuracy.

In the following, we use recall and precision as metrics to compare the performance of the three proposed methods with the OMP method from [1] under different test conditions.

C. Line Outage Identification in IEEE 39-Bus and 118-Bus Systems

1) Noise-Free Condition: We first test the performance of the methods under the perfect condition, that is, during simulation, the only perturbation is caused by line outage, other than that, the system remains at a steady state. Figure I shows the signal under this condition. Table I shows the averaged results of 10 runs per line. Results are averaged across 10 different realizations per line of the test sets..

We can see from Table I that all of the three methods we proposed outperform OMP. The advantage is more significant for larger systems. For example in the 118-bus system, the recalls of RF, LR and GCN all achieve 99%, which are superior than OMP by 13%. Among these three methods, RF and GCN are more competitive and outperform LR.

TABLE I: Noise-Free Condition

Method	System	Recall	Precision
RF	39	0.99	0.99
	118	0.99	0.99
LR	39	0.98	0.92
	118	0.99	0.984
GCN	39	0.99	0.98
	118	0.99	0.98
OMP	39	0.942	0.916
	118	0.86	0. 92

2) Noisy Conditions: This part considers the situation when measurement data are perturbed by noises. In [1] and other previous work, fixed loads plus small perturbations are considered, which account for small variations in the bus loads and can be modeled as a zero-mean Gaussian vector with a known covariance matrix. In this paper, we experiment on two cases: case 1), injecting dynamic loads at buses that cause larger load variations, and case 2), adding Gaussian noise (with dynamic load settings) to account for small load variations.

Specifically, we perturb the system during simulation by changing the powers of a subset of PQ load buses during time $25s < t < 50s$. This situation causes the measurement at each bus to exhibit extra oscillations besides the drastic changes due to line outage, thus add difficulty for the algorithms to classify line outage case. Fig. 4.(b) shows the measurements at buses with dynamic loads. Fig. 4.(c) shows the measurements with both dynamic loads and additional Gaussian noise. This is the result of adding Gaussian noise ϵ_t to measurements θ'_t at all buses, i.e., $\tilde{\theta}'_t = \theta'_t + \epsilon_t$, $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ is the noise term, with $\sigma = \frac{\alpha}{T}\sum_{t=1}^{T}\theta'_t$. Here we set $\alpha = 0.02$. We test the performance of the proposed methods and OMP under such situations. The results are shown in Table II.

From Table II we can see that for Fig. 4.(b), our methods outperform OMP by up to 12% in recall. RF and GCN are most robust to noise and can still achieve quite satisfying results from perturbed signals. LR shows slight degradation but can still achieve at least 95% in recall. GCN achieves the best results for both the 39-bus system and 118-bus system. For Fig. 4.(c), the three methods are robust to noise for small systems, the performances are only slightly degraded

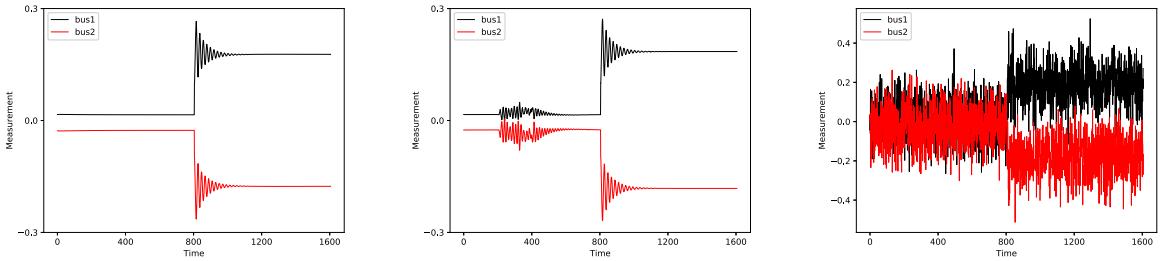


Fig. 4: Measurements θ'_1 and θ'_2 when line(1,2) is out in the 39-bus system. (a) Noise-free condition; (b) Noisy case 1, with dynamic loads; (c) Noisy case 2, with both dynamic loads and additional Gaussian noise.

TABLE II: Noisy Conditions.

Method	System	Noisy case 1		Noisy case 2	
		Recall	Precision	Recall	Precision
RF	39	0.98	0.98	0.97	0.98
	118	0.974	0.98	0.97	0.95
LR	39	0.958	0.86	0.95	0.84
	118	0.970	0.914	0.954	0.793
GCN	39	0.99	0.98	0.927	0.82
	118	0.98	0.98	0.805	0.90
OMP	39	0.9	0.85	0.5	0.46
	118	0.86	0.8	0.37	0.4

compared to the case in Fig. 4.(b). When the network is large, while RF and LR are still good but GCN is affected more. It is also noted that when noise is injected, OMP is significantly affected. This is expected, as OMP only uses the measurements at two snapshots, one before and one after the line outage. In this case, it's performance is severely impacted by the existence of noise.

3) Missing Data: This experiment deals with missing data. Missingness is a challenging problem for machine learning algorithms. We consider the task of identifying a single line outage in the 39-bus system when only a subset of buses are measured. Assume line(1,2) is the open line. The index set for the buses with missing data is denoted by B_{miss}. We distinguish

two cases with regard to the locations of the buses with missing measurements: **Case 1)** they are located close to the open line, and **Case 2)** they are located far from the open line.

Case 1): We first consider the situation when missing measurements are at the following buses, $B_{miss} = [1, 2, 3, 9, 10, 30, 39]$. Most of buses in B_{miss} are within two-hop neighborhood of the open line. Intuitively, measurements at these buses carry very important information for classification, and missing measurements can severely impact line outage identification.

In the experiment, the dataset consists of 410 inputs with 50 outage cases. Table III shows the results.

TABLE III: Results with missing data, when the missing measurements are from buses close to the open line.

Method	Recall	Precision
RF	0.84	0.98
LR	0	0
GCN	0.778	0.778
OMP	0	0

Table III shows that all three methods are impacted in different degrees when the measurements at these buses are not available. The impact is most significant to LR as it leads to complete failure of the algorithm. GCN and RF are still able to predict higher than 77% line outage cases. OMP would not work since the algorithm requires complete data as input.

Case 2): Next, we consider the situation where the missing measurements are at buses that lie far away from the open line(1,2). We randomly select a subset of such buses, indexed by $B_{miss} = [5, 6, 8, 12, 16, 21, 25, 29, 32, 34, 38]$. All these buses are located outside of at least two-hop of the open line(1,2). Under the same settings, we list the results in table IV. We can see that LR and RF are quite robust to this condition, indicating that the classifiers do not rely on measurements at buses that are not close to the open line. On the other hand, GCN shows performance degradation, although the degradation is not as severe as in the previous case where buses with missing measurements are near the open line.

TABLE IV: Results with missing data, when the missing measurements are from buses far from the open line.

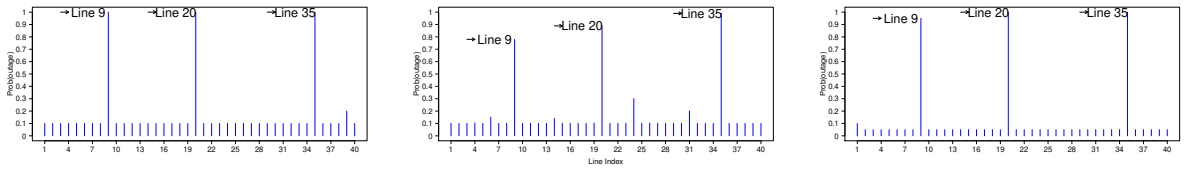
Method	Recall	Precision
RF	1	0.98
LR	1	1
GCN	0.8	0.76
OMP	0	0

D. Large System with Multiple Line Outages

We further consider three line outages in a large system, the 1354-bus system, which is part of the European transmission system. The network contains 1354 buses and 1991 branches. Due to the combinatorial complexity in enumerating all possible topologies with three line outages (there are $\binom{1991}{3}$ topologies), only 30 randomly chosen topologies are simulated. The simulation lasts for $50s$ and three lines outages simultaneously occur at time $t = 25s$ during the simulation.

TABLE V: The 1354-bus system with three line outages.

Method	Recall	Precision
RF	0.905	0.95
LR	0.919	0.95
GCN	0.93	0.99
OMP	0.24	0.3



(a) LR (b) RF (c) GCN

Fig. 5: The probability of line outage of each line from a subset of 40 lines in the 1354-bus system.

Table V shows that GCN achieves the best performance. In addition, we notice that OMP is not suitable for large systems as its recall is very low and its precision is also very low. On the other hand, the machine learning algorithms we proposed can handle such networks with satisfying performance with recalls all above 90%.

VI. Conclusion and Outlook

We have studied the line outage identification problem in power systems by using a machine learning framework. Under this framework, three learning algorithms are considered: Logistic Regression, Random Forest, and Graph Convolutional Network. They can be applied to both single line and multiple line outage identification with remarkable detection rates.

The developed line outage identification algorithms have been tested via simulation data. The IEEE 39-bus, 118-bus standard test systems and a very large 1354-bus systems are simulated in PSAT for time-domain analysis. Time series data from phasor angle measurements are used to train the algorithms. The tests on new data show that the prediction performance is satisfactory as indicated by both high precision and high recall, and the overall accuracy is comparable to methods that involve state estimation or power flow analysis.

In addition, we also addressed the situation with missing data. PMU data are often incomplete due to limited installation of PMUs in the power system. Through extensive experiments, we showed which PMU location is dispensable and which is not, and the answer is dependent on the location of the open line. In future work, we plan to furnish more rigorous analysis regarding the PMU locations and the system's ability to discriminate line topology. Additionally, real data often come with measurement errors. We showed that the proposed methods are robust to noise to a large extent.

References

[1] H. Zhu and G. B. Giannakis, "Sparse overcomplete representations for efficient identification of power line outages," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2215–2224, Nov 2012.

[2] W. B. Wu, M. X. Cheng, and B. Gou, "A hypothesis testing approach for topology error detection in power grids," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 979–985, Dec 2016.

[3] G. Andersson, P. Donalek, R. Farmer, N. Hatziargyriou, I. Kamwa, P. Kundur, N. Martins, J. Paserba, P. Pourbeik, J. Sanchez-Gasca, R. Schulz, A. Stankovic, C. Taylor, and V. Vittal, "Causes of the 2003 major grid blackouts in north america and europe, and recommended means to improve system dynamic performance," *Power Systems, IEEE Transactions on*, vol. 20, no. 4, pp. 1922–1928, Nov. 2005.

[4] "Final report on the August 14, 2003 blackout in the united states and canada: Causes and recommendations," April 5, 2004.

[5] Y. Zhao, J. Chen, A. Goldsmith, and H. V. Poor, "Identification of outages in power systems with uncertain states and optimal sensor locations," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 6, pp. 1140–1153, Dec 2014.

[6] J. E. Tate and T. J. Overbye, "Line outage detection using phasor angle measurements," *IEEE Transactions on Power Systems*, vol. 23, no. 4, pp. 1644–1652, Nov 2008.

[7] ——, "Double line outage detection using phasor angle measurements," in *2009 IEEE Power Energy Society General Meeting*, July 2009, pp. 1–5.

[8] M. Garcia, T. Catanach, S. V. Wiel, R. Bent, and E. Lawrence, "Line outage localization using phasor measurement data in transient state," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 3019–3027, July 2016.

[9] T. Banerjee, Y. C. Chen, A. D. Dominguez-Garcia, and V. V. Veeravalli, "Power system line outage detection and identification — a quickest change detection approach," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 3450–3454.

[10] J. Chen, Y. Zhao, A. Goldsmith, and H. V. Poor, "Line outage detection in power transmission networks via message passing algorithms," in *2014 48th Asilomar Conference on Signals, Systems and Computers*, Nov 2014, pp. 350–354.

[11] Y. Zhao, J. Chen, and H. V. Poor, "Efficient neural network architecture for topology identification in smart grid," in *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec 2016, pp. 811–815.

[12] C. Lee and S. J. Wright, "Using neural networks to detect line outages from PMU data," *CoRR*, vol. abs/1710.05916, 2017. [Online]. Available: http://arxiv.org/abs/1710.05916

[13] Y. Liao, Y. Weng, C. Tan, and R. Rajagopal, "Urban distribution grid line outage identification," in *2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, Oct 2016, pp. 1–8.

[14] T. Kim and S. J. Wright, "PMU placement for line outage identification via multinomial logistic regression," *IEEE Transactions on Smart Grid*, vol. 9, no. 1, pp. 122–131, Jan 2018.

[15] A. Y. Abdelaziz, S. F. Mekhamer, M. Ezzat, and E. F. El-Saadany, "Line outage detection using support vector machine (SVM) based on the phasor measurement units (pmus) technology," in *2012 IEEE Power and Energy Society General Meeting*, July 2012, pp. 1–8.

[16] M. He and J. Zhang, "A dependency graph approach for fault detection and localization towards secure smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 342–351, June 2011.

[17] M. X. Cheng, Y. Ling, and W. B. Wu, "In-band wormhole detection in wireless ad hoc networks using change point detection method," in *2016 IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–6.

[18] D. R. Cox, "The regression analysis of binary sequences," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 20, no. 2, pp. 215–242, 1958. [Online]. Available: http://www.jstor.org/stable/2983890

[19] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001. [Online]. Available: https://doi.org/10.1023/A:1010933404324

[20] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3042–3054, 2014.

[21] F. Milano, L. Vanfretti, and J. C. Morataya, "An open source power system virtual laboratory: The psat case and experience," *IEEE Transactions on Education*, vol. 51, no. 1, pp. 17–23, February 2008.

[22] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs," *IEEE transactions on signal processing*, vol. 61, no. 7, pp. 1644–1656, 2013.

[23] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[24] M. R. Leadbetter, G. Lindgren, and H. Rootzén, *Extremes and related properties of random sequences and processes*, ser. Springer Series in Statistics. MIR, 1989.

APPENDIX

A. Logit Function

In logistic regression ([18]), Y is the binary outcome, $(x_1, \ldots, x_m)$ is the set of predictor variables, and p is the probability of Y being 1.

Using the training data and the following logit link function,

$$\log_e \left(\frac{p}{1-p} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_m x_m = \boldsymbol{\beta} \cdot \mathbf{X}$$

we can estimate the coefficients $\hat{\boldsymbol{\beta}}$.

For the prediction of new data at $\mathbf{X_h}$, using the following can compute the probability of $Y_h = 1$,

$$\mathbb{P}(Y_h = 1) = \frac{1}{1 + e^{-\hat{\boldsymbol{\beta}} \cdot \mathbf{X_h}}}$$

B. Graph Fourier Transform

Graph Fourier transform extends the idea of discrete Fourier transform to vectors defined on a graph. We first provide a graph interpretation of DFT and show how it can be extended to GFT.

The discrete Fourier transform of a vector $\boldsymbol{s}$ is given by

$$\widehat{s}_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} s_n e^{-i \frac{2\pi}{N} kn}, \ k = 0, \cdots, N-1$$

where $\widehat{s}_k$ is the element of vector $\widehat{\boldsymbol{s}}$.

The discrete Fourier transform of a vector can be written as the multiplication of a matrix $\mathbf{DFT}$ with a vector $\boldsymbol{s}$,

$$\widehat{\boldsymbol{s}} = \mathbf{DFT}\boldsymbol{s} \tag{5}$$

Let $\omega = e^{-i2\pi/N}$. The $\mathbf{DFT}$ matrix can then be defined as follows:

$$\mathbf{DFT} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \omega & \cdots & \omega^{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{k} & \cdots & \omega^{k(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \cdots & \omega^{(N-1)(N-1)} \end{bmatrix}$$

The row vectors of $\mathbf{DFT}$ are $\frac{1}{\sqrt{N}}[1, \omega^{k}, \cdots, \omega^{k(N-1)}],\ k = 0, \cdots, N-1$, which are the spectral components.

To give DFT operation a graph interpretation, we can consider vector $\boldsymbol{s}$ as a signal defined on a directed cycle graph $0 \to 1 \to \ldots n-1 \to 0$ ([22], [23]). The adjacency matrix of the cycle graph, in which element $A_{i,j}$ represents the edge from j to i, is the following matrix,

$$\mathbf{A}_c = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

The eigen-decomposition of $\mathbf{A}_c$ gives

$$\begin{aligned} \mathbf{A}_c &= \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1} \\ &= \mathbf{DFT}^{-1} \begin{pmatrix} e^{-i\frac{2\pi}{N}\cdot 0} & & \\ & \ddots & \\ & & e^{-i\frac{2\pi\cdot(N-1)}{N}} \end{pmatrix} \mathbf{DFT} \end{aligned} \tag{6}$$

Columns of $\mathbf{V}$ are eigenvectors of $\mathbf{A}_c$, and $\boldsymbol{\Lambda}$ is a diagonal matrix of the eigenvalues of $\mathbf{A}_c$. Since the frequencies $\frac{2\pi k}{N}$ of a DFT operation correspond to the eigenvalues of $\mathbf{A}_c$, we can view the eigenvalues as a representation of frequencies. Analogously, $\mathbf{V}^{-1}$ is considered playing the role of the $\mathbf{DFT}$ matrix in the context of Fourier transform of graph signals.

The connection between DFT and circulant adjacency matrix of time signal leads to the natural definition of Fourier transform operated on graph signals. We now extend from the directed cycle graph to general graphs, which only requires to replace $\mathbf{A}_c$ with the normalized adjacency matrix $\tilde{A}$.

Let $\tilde{A}$ be the normalized adjacency matrix of the power grid network. The eigen-decomposition of $\tilde{A}$ gives

$$\tilde{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1} \tag{7}$$

Here $\mathbf{\Lambda} = \operatorname{diag}[\lambda_0 \cdots \lambda_{N-1}]$ is a complete set of eigenvalues. $\mathbf{V} = [\boldsymbol{v}_0, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_{N-1}]$ is an $N \times N$ matrix with $\boldsymbol{v}_i$ denoting the eigenvector corresponding to λ_i. The eigenvectors are the graph spectral components.

Now we wish to define the Graph Fourier Transform of a signal defined on a graph analogously by replacing matrix $\mathbf{DFT}$ with its graph counterpart $\mathbf{GFT}$. We define matrix $\mathbf{GFT} = \mathbf{V}^{-1}$. Analogous to DFT in (5), the graph Fourier transformation of a vector $\boldsymbol{s}$ is also given as a matrix-vector multiplication:

$$\widehat{\boldsymbol{s}} = \mathbf{GFT}\boldsymbol{s} = \mathbf{V}^{-1}\boldsymbol{s} = [\widehat{s}_0, \widehat{s}_1, \cdots, \widehat{s}_{N-1}]^\top \tag{8}$$

where $\widehat{s}_k$ is the Fourier coefficient corresponds to λ_k.

C. Gumbel Distribution Critical Value

FFT results in a sequence of $\{X_k\}$, with $k = 0, \ldots, n-1$. Let $M_n = \max_{1 \le k \le n} |X_k|^2$.

M_n follows Gumbel distribution asymptotically. It has been proved in [24] that there exists two sequences (a_n) and (b_n) such that

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{M_n - a_n}{b_n} \le t\Big) = e^{-e^{-t}}.$$

With the level of significance α (e.g., $\alpha = 0.05$), we have $1 - \alpha = e^{-e^{-t}}$. Solving it for t, we get critical value $t = -\log(-\log(1-\alpha))$. Then we can select the dominant frequencies using this critical value.