LEARNING TO INFER: A NEW VARIATIONAL INFERENCE APPROACH FOR POWER GRID TOPOLOGY IDENTIFICATION

Yue Zhao, Jianshu Chen† and H. Vincent Poor‡*

*Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, NY 11794
†Microsoft Research, Redmond, WA 98052
‡Department of Electrical Engineering, Princeton University, Princeton, NJ 08544
Email: yue.zhao.2@stonybrook.edu, jianshuc@microsoft.com, poor@princeton.edu

ABSTRACT

Identifying arbitrary topologies of power networks is a computationally hard problem due to the number of hypotheses that grows exponentially with the network size. A new variational inference approach is developed for efficient marginal inference of every line status in the network. Optimizing the variational model is transformed to and solved as a discriminative learning problem. A major advantage of the developed learning based approach is that the labeled data used for learning can be generated in an arbitrarily large amount at very little cost. As a result, the power of offline training is fully exploited to offer effective real-time topology identification. The proposed methods are evaluated in the IEEE 30-bus system. With relatively simple variational models and only an undercomplete measurement set, the proposed method already achieves very good performance in identifying arbitrary power network topologies.

***Index Terms*—** Power grid topology identification, line outage detection, machine learning, variational inference

1. INTRODUCTION

Understanding the network status in abnormal system conditions is crucial for preventing blackouts in power networks. Network component failures (e.g. line outages), if unattended, can quickly escalate to cascading failures that grow out of control of the system operator. When line failures happen, the power network topology changes instantly, newly stressed areas can unexpectedly emerge, and subsequent failures may be triggered. Real time network topology identification is thus essential to all network control decisions for mitigating failures.

Topology identification is however a very challenging problem, especially when unknown line statuses in the network accumulate as in scenarios that cause large-scale blackouts [1]. The number of possible topologies grows exponentially with the number of unknown line statuses, making topology identification fundamentally hard. In practice, time pressure and incomplete information can make this problem even harder. Prior works on line outage detection have primarily focused on scenarios with a small number of line failures. Exhaustive search methods have been developed in [2], [3], [4] and [5] based on hypothesis testing, and in [6] and [7] based on logistic regression. To overcome the prohibitive computational complexity of exhaustive search methods, [8] has exploited the sparsity of outage patterns with overcomplete observations to identify sparse multi-line outages. Recently, a graphical model based approach has been developed for identifying arbitrary network topologies without assuming sparsity of line outages [9]. However, ensuring robustness of message passing algorithms for large-scale networks remains a challenging task.

In this paper, we develop a new "learning-to-infer" approach for identifying arbitrary topologies of power networks. We start with a probabilistic model in which the variables in power networks are modeled in a Bayesian framework. We observe that exhaustively expressing the mapping between the physical quantities in power networks (e.g. nodal power injections and voltages) and the network topology is computationally hard due to the exponentially large number of possible topologies. To overcome this hardness, we develop a variational inference framework, in which we approximate this mapping using models that allow computationally easy marginal inference of line statuses. Furthermore, with a Monte Carlo approach, optimizing the variational model is transformed to a discriminative learning problem. In particular, data samples of network topology, power injections, and measurements in the network can be easily generated according to a generative model of these quantities. With these data, discriminative models for predicting the network topology based on the available measurements and the power injections are trained and tested.

A major strength of the proposed approach is that the *labeled data set* for training the variational model can be generated in an arbitrarily large amount, at very little cost. As

This research was supported in part by the National Science Foundation under Grant ECCS-1549881.

such, we can fully exploit the benefit of offline model training in order to get accurate online topology identification performance. The proposed approach is also not restricted to specific learning methods, but can exploit any powerful models such as deep neural networks. The developed learning-to-infer method is evaluated in the IEEE 30-bus network [10] for identifying topologies with an *arbitrary* number of line outages using only *undercomplete* measurements. It is demonstrated that, even with relatively simple variational models, the performance is surprisingly good for this very challenging task.

2. SYSTEM MODEL

We consider a power network with N buses, and a baseline topology with L lines. We denote the incidence matrix of the baseline topology by $M \in \{-1, 0, 1\}^{N\times L}$. We use a binary variable s_l to denote the status of a line l, with $s_l = 1$ for a connected line l, and 0 otherwise. The actual topology of the network can then be represented by $\boldsymbol{s} = [s_1, \ldots, s_L]^T$. In this paper, we employ the DC power flow model for brevity [11]. We note that the developed methodology can be directly extended to the AC power flow model. We denote the power injections and voltage phase angles at all the buses by $\boldsymbol{P} \in \mathbb{R}^N$ and $\boldsymbol{\theta} \in \mathbb{R}^N$, respectively. Based on the DC power flow model, we have

$$\boldsymbol{P} = MS\Gamma M^T\boldsymbol{\theta}, \tag{1}$$

where $S = \mathrm{diag}(s_1, \ldots, s_L)$, $\Gamma = \mathrm{diag}(\frac{1}{x_1}, \ldots, \frac{1}{x_L})$, and x_l is the reactance of line l.

We focus on identifying the network topology $\boldsymbol{s}$ based on real time measurements of $\boldsymbol{\theta}$ provided by phasor measurement units (PMUs) located at a subset of the buses $\mathcal{M}$, as well as knowledge of $\boldsymbol{P}$. We model the PMU measurements as

$$\boldsymbol{y} = \boldsymbol{\theta}_{\mathcal{M}} + \boldsymbol{v}, \tag{2}$$

where $\boldsymbol{\theta}_{\mathcal{M}}$ is formed by entries of $\boldsymbol{\theta}$ from buses in $\mathcal{M}$, and $\boldsymbol{v} \sim N(0, \sigma^2 I)$ contains the measurement noise.

We formulate the topology identification problem as a probabilistic inference problem. First, we model $\boldsymbol{s}, \boldsymbol{P}$ and $\boldsymbol{y}$ with a joint probability distribution that can be expressed as $p(\boldsymbol{s}, \boldsymbol{P}, \boldsymbol{y}) = p(\boldsymbol{s}, \boldsymbol{P})p(\boldsymbol{y}|\boldsymbol{s}, \boldsymbol{P})$. Note that $p(\boldsymbol{y}|\boldsymbol{s}, \boldsymbol{P})$ is fully determined by the power flow model (1) and the observation model (2). We further assume that the power injections are in static or quasi-steady state, so that $\boldsymbol{P}$ is known when performing topology identification. Our objective is to infer the topology of the power grid, characterized by $\boldsymbol{s}$, given the observed $\boldsymbol{y}$ and the prior knowledge $\boldsymbol{P}$. This can be achieved by finding the posterior probability $p(\boldsymbol{s}|\boldsymbol{y}, \boldsymbol{P})$. However, as there are in total 2^L possibilities for $\boldsymbol{s}$, computing, or even expressing the probability $p(\boldsymbol{s}|\boldsymbol{y}, \boldsymbol{P})$ has an exponential complexity. As a result, it is in general hard to perform inference tasks such as finding the line status marginals $p(s_l|\boldsymbol{y}, \boldsymbol{P})$, or finding the topology $\boldsymbol{s}$ with the maximum a-posteriori probability.

3. A LEARNING-TO-INFER VARIATIONAL INFERENCE FRAMEWORK

Due to the aforementioned challenges in the exact computation of the posterior $p(\boldsymbol{s}|\boldsymbol{y}, \boldsymbol{P})$, we proceed to develop an approximate inference method for the marginal posterior $p(s_l|\boldsymbol{y}, \boldsymbol{P}), l = 1, \ldots, L$, by a variational method. We seek to find a variational distribution $q(\boldsymbol{s}|\boldsymbol{y}, \boldsymbol{P})$ to approximate the original $p(\boldsymbol{s}|\boldsymbol{y}, \boldsymbol{P})$ as much as possible by minimizing the Kullback-Leibler divergence $D(p\|q)$, i.e., finding the M-projection [12] of p. We require this variational distribution $q(\boldsymbol{s}|\boldsymbol{y}, \boldsymbol{P})$ to satisfy the following:

- It has sufficient expressive power to represent complicated functions so that our approximation to $p(s_l|\boldsymbol{y}, \boldsymbol{P})$ can be made sufficiently precise.
- It is easy to compute $q(s_l|\boldsymbol{y}, \boldsymbol{P})$, and we can use it to infer s_l based on the observed $\boldsymbol{y}$ and $\boldsymbol{P}$ with low computation complexity.

In practice, we restrict $q(\boldsymbol{s}|\boldsymbol{y}, \boldsymbol{P})$ to have special parametric forms, $q_{\boldsymbol{\beta}}(\boldsymbol{s}|\boldsymbol{y}, \boldsymbol{P})$, that are easy to compute, where $\boldsymbol{\beta}$ is a vector of model parameters. Minimizing $D(p\|q)$ is then equivalent to finding the $q_{\boldsymbol{\beta}}(\boldsymbol{s}|\boldsymbol{y}, \boldsymbol{P})$ that solves the following optimization problem:

$$\max_{\boldsymbol{\beta}} \mathbb{E}_p[\log q_{\boldsymbol{\beta}}(\boldsymbol{s}|\boldsymbol{y}, \boldsymbol{P})], \tag{3}$$

where the expectation is taken with respect to the true distribution p. Furthermore, as $\mathbb{E}_p[\log q_{\boldsymbol{\beta}}(\boldsymbol{s}|\boldsymbol{y}, \boldsymbol{P})]$ can be hard to compute, we can approximate it by the empirical mean of $\log q_{\boldsymbol{\beta}}(\boldsymbol{s}|\boldsymbol{y}, \boldsymbol{P})$ over a large number of Monte Carlo samples generated according to the joint probability $p(\boldsymbol{s}, \boldsymbol{P}, \boldsymbol{y})$, denoted by $\{\boldsymbol{s}^t, \boldsymbol{P}^t, \boldsymbol{y}^t, t = 1, \ldots, T\}$. This yields,

$$\max_{\boldsymbol{\beta}} \frac{1}{T}\sum_{t=1}^{T} \log q_{\boldsymbol{\beta}}(\boldsymbol{s}^t|\boldsymbol{y}^t, \boldsymbol{P}^t)]. \tag{4}$$

With the generated data set $\{\boldsymbol{s}^t, \boldsymbol{P}^t, \boldsymbol{y}^t\}$, (4) can be efficiently solved, and the optimal model parameters $\boldsymbol{\beta}$ approaches those for the desired M-projection as $T \to \infty$.

In fact, the problem (4) can be viewed as an empirical risk minimization problem in machine learning [13], which trains a discriminative model $q_{\boldsymbol{\beta}}(\boldsymbol{s}|\boldsymbol{y}, \boldsymbol{P})]$ with a data set $\{\boldsymbol{s}^t, \boldsymbol{P}^t, \boldsymbol{y}^t\}$ generated from a generative model $p(\boldsymbol{s}, \boldsymbol{P}, \boldsymbol{y})$. As such, the *inference* problem of topology identification is cast as a *learning* problem, as we train a model $q_{\boldsymbol{\beta}}(\boldsymbol{s}|\boldsymbol{y}, \boldsymbol{P})$ so that it can predict the line status $\boldsymbol{s}$ given any newly observed measurements $\boldsymbol{y}$ and knowledge of $\boldsymbol{P}$. In particular, we would like to compute the line status marginals $q(s_l|\boldsymbol{y}, \boldsymbol{P})$

and, eventually, identify whether each line l is connected or not. Thus, we can employ multi-label classifiers where each binary line status corresponds to one label.

One great advantage of this learning-to-infer approach is that we can generate *labeled data* very efficiently. Specifically, we can efficiently sample from the generative model $p(\boldsymbol{s}, \boldsymbol{P}, \boldsymbol{y}) = p(\boldsymbol{s}, \boldsymbol{P})p(\boldsymbol{y}|\boldsymbol{s}, \boldsymbol{P})$ as long as we have some prior $p(\boldsymbol{s}, \boldsymbol{P})$ that is easy to sample from. While historical data and expert knowledge would surely help in forming such priors, using simple uninformative priors can already suffice as will be shown later in the numerical examples. As a result, we can obtain *an arbitrarily large set of data with very little cost* to train the discriminative model. This is quite different from the typical situations encountered in machine learning problems, where obtaining a large amount of labeled data is usually expensive as it requires extensive human annotation effort.

Another advantage of our proposed method is that, once the approximate posterior probability q is learned, it can be deployed to infer the power grid topology in real-time as the computation complexity of $q(s_l|\boldsymbol{y}, \boldsymbol{P})$ is very low. This is especially important in monitoring large-scale power grids in real time, because, although training q could take a reasonably amount of time, the inference speed is very fast. Therefore, the learned predictor q can be used in real-time with low-cost hardware.

4. NUMERICAL EVALUATION

We evaluate the proposed learning-to-infer approach for topology identification with the IEEE 30 bus system as the baseline topology (cf. Figure 1). There are 41 lines in total. As opposed to considering only sparse line outages as in existing works, we allow *any number* of line outages, and investigate whether the learned discriminative classifiers can successfully recover the topologies.

4.1. Data Set Generation

To generate a data set $\{\boldsymbol{s}^t, \boldsymbol{P}^t, \boldsymbol{y}^t\}$, we assume the prior $p(\boldsymbol{s}, \boldsymbol{P})$ factors as $p(\boldsymbol{s})p(\boldsymbol{P})$. We generate the line statuses $\{s_l\}$ using independent and identically distributed (IID) Bernoulli random variables with $p(s_l = 1) = 0.6$. We do not consider disconnected networks in this study, and exclude the line status samples if they lead to disconnected networks. As a result, there are three lines (9-11, 12-13, 25-26) that are always connected in these topologies, and the dimension of the vector $\boldsymbol{s}$ reduces to 38.

We would like our predictor to be able to identify the topology for *generic values of power injections* as opposed to fixed ones. Accordingly, we generate $\boldsymbol{P}$ using the following procedure: We first generate bus voltage phase angles $\boldsymbol{\theta}$ as IID uniform random variables in $[0, 0.2\pi]$, and then compute $\boldsymbol{P}$ according to (1) under the baseline topology. Lastly,

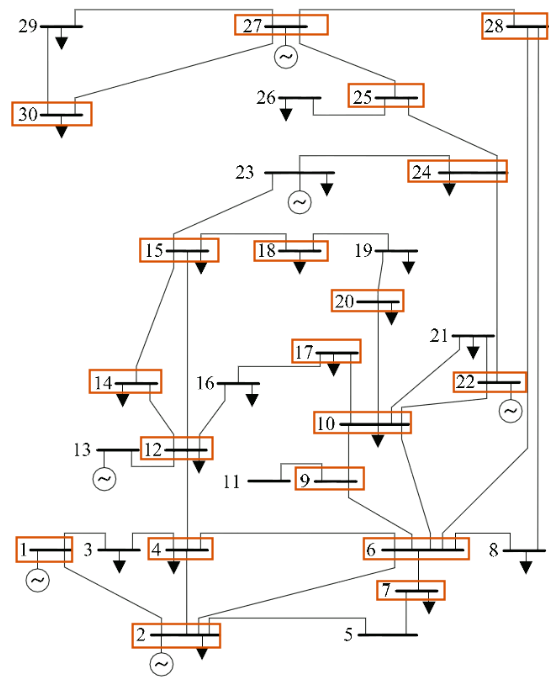

Fig. 1. IEEE 30-bus system, and the PMU locations.

with each pair of generated $\boldsymbol{s}^t$ and $\boldsymbol{P}^t$, we generate IID phase angle measurement noise with a standard deviation of 0.01 degree, the state-of-the-art PMU accuracy [14].

In this study, we generate a total of $300K$ data samples for training and testing procedures. In comparison, the total number of connected topologies of the IEEE 30 bus system is on the order of $2^{38} = 2.75 \times 10^{11}$. We note that over 99% of the generated $300K$ connected topologies are distinct from each other. The *average number of disconnected lines* relative to the baseline topology is 7.8, which is significantly higher than those typically assumed in sparse line outage studies.

Finally, we choose *19 among the 30* bus voltage phase angles as the measurement set, as depicted in Figure 1. We note that, while the selected measurement locations will be shown to lead to very good identification performance, optimizing these locations more systematically is left as an interesting topic for future work. We will show that even with such a severely undercomplete set of measurements, we can already identify very well the network topology among an exponentially large number of hypotheses.

4.2. Identifying $\boldsymbol{s}$ with Multi-Label Classifiers

We employ a binary relevance method (i.e., independently training one classifier for each label s_l). For identifying $\boldsymbol{s}$ given $\boldsymbol{y}$ and $\boldsymbol{P}$, we train and compare three classifiers with C4.5 decision trees [15], multilayer perceptrons [16] and logistic regression, respectively. In particular, with the multilayer perceptron, we employ one hidden layer with 24 hid-

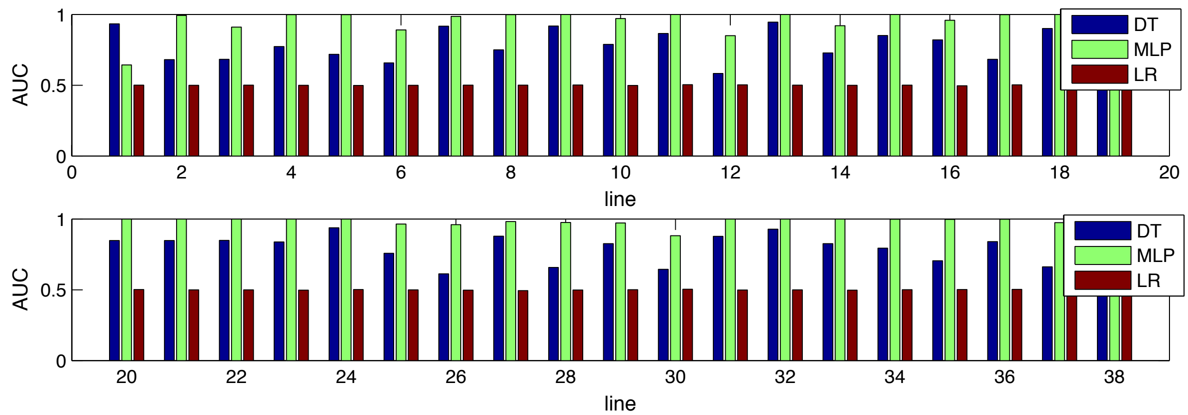

Fig. 2. Areas under the ROC for line status identification using a C4.5 Decision Tree (DT), a Multilayer Perceptron (MLP), and Logistic Regression (LR).

den units. We use the MEKA open source toolkit for implementing these learning methods [17]. Among the $300K$ data points, we use $\frac{2}{3}$ of them as the training data, and $\frac{1}{3}$ as the testing data.

4.3. Evaluation Results

For identifying each line status, we obtained the receiver operating characteristic (ROC) with each of the three classifiers, and computed the area under the curve (AUC). All the AUCs are plotted in Figure 2. *The average AUC for the 38 lines is 0.792 with C4.5 decision trees, 0.969 with multilayer perceptrons, and 0.5 with logistic regression.*

We observe that the multilayer perceptron achieves surprisingly good performance in identifying the network topology, which has been a hard open problem not known to be tractable in the literature. Moreover, this is achieved using only 19 (out of 30) voltage phase angle measurements. On a laptop with an Intel Core i7 3.1-GHz CPU and 8 GB of RAM, with $200K$ data points, it takes about 16 hours to train the multilayer perceptron model, one hour to train the decision tree model, and 12 minutes to train the logistic regression model. Moreover, it takes about only 38 seconds to test the identification performance on $100K$ data samples (which is 0.38 milliseconds per sample) with the multilayer perceptron, and 18 seconds with the two other models. The extremely fast testing speed demonstrates that the proposed approach applies very well to real time tasks, such as failure identification during cascading failures. In comparison to the multilayer perceptron, we observe that a) using a C4.5 decision tree provides a lower but reasonably good performance with a shorter training time, and b) a classifier trained by logistic regression provides essentially no predictive power (AUC $= 0.5$) for topology identification.

We would like to further emphasize that the topologies and the power injections used to train the predictor q are different from the ones in the test set. This is of particular interest because it means that our learned predictor q is able to *generalize* well on the *unseen* test topologies and power injections based on its knowledge learned from the training data.

It is also worth noting that we have generated the training and testing data set with uniformly random voltage phase angles, and hence considerably variable power injections. In practice, there is often more informative prior knowledge about the power injections based on historical data and load forecasts. With such information, the model can be trained with much less variable samples of power injections, and the identification performance can be further improved significantly. For example, if we fix the power injections to a set of typical values (i.e., fixing the prior $p(\boldsymbol{P})$ to be deterministic), the average AUC increases to above 0.97 even with the simpler C4.5 decision tree classfier.

5. CONCLUSION

We have developed new learning-to-infer variational inference methods for topology identification of power grids. The computational complexity due to the exponentially large number of hypotheses is overcome by efficient marginal inference with the variational model. Optimization of the variational model is transformed to and solved as a discriminative learning problem. The developed methods have the major advantage that the labeled data set can be generated in an arbitrarily large amount with very little cost. As a result, the variational model can always be trained with sufficient data, so that excellent online topology identification performance can be achieved. We have evaluated the proposed methods with the IEEE 30-bus system employing C4.5 decision trees, multilayer perceptrons, and logistic regression in discriminative learning. With the multilayer perceptron, it has been demonstrated that arbitrary network topologies can be identified with very good performance using only 19 (among 30) voltage phase angle measurements.

6. REFERENCES

[1] *US-Canada Power System Outage Task Force, Final Report on the August 14, 2003 Blackout in the United States and Canada*, 2004.

[2] 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.

[3] J. E. Tate and T. J. Overbye, "Double line outage detection using phasor angle measurements," in *Proc. IEEE Power and Energy Society General Meeting*, July 2009.

[4] 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.

[5] Y. Zhao, A. Goldsmith, and H. V. Poor, "On PMU location selection for line outage detection in wide-area transmission networks," in *Proc. IEEE Power and Energy Society General Meeting*, July 2012, pp. 1–8.

[6] T. Kim and S. J. Wright, "PMU placement for line outage identification via multiclass logistic regression," *arXiv preprint arXiv:1409.3832*, 2014.

[7] 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, to appear*.

[8] 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.

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

[10] *Power Systems Test Case Archive, University of Washington Electrical Engineering*.

[11] J. D. Glover, M. Sarma, and T. Overbye, *Power System Analysis & Design*, Cengage Learning, 2011.

[12] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*, MIT press, 2009.

[13] V. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.

[14] A. von Meier, D. Culler, A. McEachern, and R. Arghandeh, "Micro-synchrophasors for distribution systems," in *Proc. IEEE Innovative Smart Grid Technologies (ISGT)*, 2013.

[15] J. R. Quinlan, *C4. 5: Programs for Machine Learning*, Elsevier, 2014.

[16] S. Haykin, *Neural Networks: A Comprehensive Foundation*, MacMillan Publishing Company, 1994.

[17] *MEKA: A Multi-label Extension to WEKA*.