Fingerprint Discovery for Transformer Health Prognostics from Micro-Phasor Measurements

Jose Cadena [1] **Priyadip Ray** [1] **Emma Stewart** [1]

Abstract

A key component in the transfer of electric power from the generator to the consumer is the power transformer. Each distribution feeder may have hundreds of devices spread throughout a large geographical area. Transformer failures are a key indicator of grid resiliency and reliability. The availability of large quantities of high-granularity data obtained from micro-phasor measurement units (μPMUs) provides a unique opportunity for applying machine learning (ML) techniques for transformer health diagnostics and prognostics. Here, we propose a Bayesian non-parametric model for uncovering "temporal signatures" related to transformer health, which may be utilized for developing risk stratification systems. We provide results on grid data from Riverside, CA to demonstrate the efficacy of our proposed approach.

1. Introduction

Transformers are responsible for the majority of transfer of power across the power grid; consequently, transformer failures account for a dominant fraction of equipment issues and are a key limitation to grid resiliency and reliability. Current power distribution system diagnostics generally focus on reaction to events, either outages or power quality events. However, longer-term degradation and indicators of impending failure are not mainstream operational applications at present. The advent of distribution-level phasor measuring systems, i.e., μPMUs (Stewart et al., 2014), which allow data collection at a much higher granularity than previously possible, has opened up new possibilities of employing ML for improved transformer health monitoring and prognostics. Armed with an effective failure prediction, utilities could perform preventative maintenance, anticipatory unloading, or repair, all before a catastrophic failure—improving system reliability and resilience, saving the utility time and money, and improving customer quality of service.

In the last few years, machine learning and data-driven approaches have been widely applied in smart grid systems (Ali et al., 2016; Shyam et al., 2015; Zhang et al., 2010; Diamantoulakis et al., 2015). In particular, μPMU data has been used for monitoring, protection and control of distribution systems assets, and analyzing switching effects of capacitor banks (Stewart et al., 2017; Shahsavari et al., 2017). However, transformer health prognostics from micro-phasor measurements pose different challenges. First, obtaining labelled data of "risky" or erratic transformer behavior is difficult; thus, unsupervised learning approaches are needed to identify different patterns of activity in the grid—e.g., normal activity, sudden drop, etc. Second μPMU data is finely sampled and highly correlated, so techniques for discovering risky transformer behavior need to be scalable.

1.1. Contributions

To address the above challenges, we propose the task of *fingerprint discovery* for transformer health prognostics. We model μPMU time series data as a probabilistic mixture of K recurring temporal signatures or fingerprints, which correspond to different operating states of a transformer.

We propose an unsupervised Bayesian non-parametric approach for automatic segmentation and clustering of μPMU data. Specifically, we model a time series using a Hierarchical Dirichlet Process (HDP) (Teh et al., 2006) with continuity constraints. The time series is first partitioned into non-overlapping segments according to a distance-dependent Chinese Restaurant Process (ddCRP) (Blei & Frazier, 2011), ensuring temporal continuity. Then, each segment is associated with one out of K signatures—where K is inferred—according to a standard Chinese Restaurant Process. Our approach is simple to interpret, scalable, and can automatically discover both the number of non-overlapping change points in the time series, as well as the number of fingerprints that best describe the observed data.

[1]Engineering Directorate, Lawrence Livermore National Laboratory, Livermore, CA, USA.
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 and was supported by the LLNL-LDRD Program under Project No. 19-ERD-009. Correspondence to: Jose Cadena <cadenapico1@llnl.gov>.

Proceedings of the 36^{th} International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019.

2. Fingerprint discovery from μPMUs

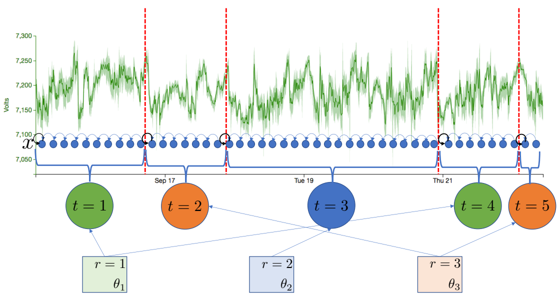

Figure 1. Finding temporal fingerprints. Observations (blue dots) are clustered into realizations of one out of K fingerprints (3 fingerprints in the figure—green, blue, orange). Temporal segments 1 and 4 are realizations of the same fingerprint, and all observations within the segment are generated from the same distribution with parameters θ_1.

A μPMU is a high-precision sensor designed for making synchronized, time-stamped measurement of voltage magnitude and phase angle at the power distribution level. We assume that the observed μPMU voltage measurements $\mathbf{X} = \{x_1, \ldots, x_T\}$ are a concatenation of K different temporal *signatures* or *fingerprints*, $r = 1, \ldots, K$. These signatures could correspond, for instance, to different types of failures in the power grid, as well as normal activity. Our goal is to uncover these different temporal signatures from observed data; later, using domain expert knowledge, these temporal signatures may potentially be used for risk stratification. A temporal signature r is characterized by a set of parameters θ_r. Each temporal signature may appear multiple times in different parts of the time series and have a different duration. However, we constrain each realization of the signature to be contiguous and non-overlapping with adjacent signatures.

Another interpretation, which is convenient for describing the model, is that the time series can be partitioned into contiguous segments, and these segments themselves can be grouped into a discrete set of K classes (i.e., the signatures). In order to cluster observations temporally, we employ a distance-dependent Chinese Restaurant (ddCRP) prior. Briefly, the ddCRP is a generative Bayesian non-parametric model for clustering data. The generative model may be described via a restaurant analogy, where there is a restaurant with an infinite number of tables, and there is a set of customers $\{1, \ldots, n\}$ who walk into the restaurant and sit at a table one at a time. Additionally, we are given a distance matrix D, such that $D_{i,j}$ is the distance between a pair of customers. On entering the restaurant, each customer i either chooses to sit at a table by himself or sit with an existing customer. With probability proportional to a parameter α, the customer chooses an empty table and orders a dish θ for the table. Alternatively, i joins an existing customer j with probability proportional to a decreasing function of their distance, called the *decay*. Let $c_i = j$ be i's customer choice (note, $c_i = i$ if i sits by himself). Then, the customer assignments are drawn as follows:

$$p(c_i = j|\text{decay}, D, \alpha) \propto \begin{cases} \text{decay}(D_{i,j}) & i \neq j \\ \alpha & i = j \end{cases} \tag{1}$$

In our application, $D_{i,j}$ is simply the distance in the time series: $D_{i,j} = |i - j|$. The decay function is defined as

$$\text{decay}(d) = \begin{cases} 1 & d \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

In other words, we only allow i to sit next to an adjacent customer or to start a new table. This way, temporal continuity is enforced. Note that the customer assignments c_i induce a clustering of the data. We use t_i to refer to the cluster or segment of observation x_i. Each one of these clusters corresponds to one realization of a temporal signature in our model. Now, we model clusters of segments using a standard Chinese Restaurant Process with dispersion parameter λ. Each segment t obtained before is now a customer who sits at a table r_t with probability proportional to the number of customers already there, or at a new table with probability γ. Finally, given the region r, observations assigned to that region are i.i.d. samples from a distribution f parameterized by θ_r—e.g., a Gaussian distribution. The full generative model is described below and illustrated in Figure 2.

1. For each observation x_i, sample a customer assignment $c_i \sim \text{ddCRP}(\text{decay}, D, \alpha)$, which gives us a set of tables $t = 1, \ldots, M$. Let t_i be the table of x_i
2. For each table t, sample a region assignment $r_t \sim \text{CRP}(\lambda)$. Let $z_i = r_{t_i}$ be the region of observation x_i
3. For each region $r = 1, \ldots, K$, sample the region parameters $\theta_r = (\mu_r, \sigma_r^2)$
4. For each observation, x_i, sample $x_i|z_i \sim \mathcal{N}(\mu_{z_i}, \sigma_{z_i}^2)$

2.1. Inference

We perform inference via Gibbs sampling (Blei & Frazier, 2011; Teh et al., 2006). Let c_{i-} denote the current customer assignment for all customers except i. The **new customer assignment** c_i for customer i is sampled from

$$p(c_i|c_{i-}, r, \text{decay}, \alpha, D, \mathbf{X}, \theta) \propto \\ p(c_i|\text{decay}, \alpha, D) p(\mathbf{X}|c_i, c_{i-}, r, \theta).$$

The last term in the equation above gives the likelihood of the data under the new assignment. In the case, that $c_i \neq i$—i.e., i's segment joins an existing fingerprint, this likelihood can be written as a product of the likelihoods for each fingerprint:

$$p(\mathbf{X}|c_i, c_{i^-}, r, \theta) = p(\mathbf{X}|t, r, \theta) \quad (2)$$

$$= \prod_{k=1}^{K} p(\{x_j | r_{t_j} = k\} | \theta_k). \quad (3)$$

Otherwise, we have to sample a new fingerprint for i, and the likelihood of the data is given by integrating out all the possible fingerprint assignments, including a new fingerprint assignment $K+1$:

$$\sum_{k=1}^{K+1} p(r_{t_i} = k | r_{t_i^-}, \lambda) p(\mathbf{X}|c_i = i, c_{i^-}, r_{t_i} = k, r_{t_i^-}, \theta)$$

Sampling the **new fingerprint assignment** r_t for each segment t simply requires draws from:

$$p(r_t | r_{t^-}, \lambda, \mathbf{X}, \theta) \propto p(r_t | r_{t^-}, \lambda) p(\mathbf{X}|r_t, r_{t^-}, \theta),$$

where the likelihood of the data is the same as in 3.

3. Experiments

3.1. Riverside Public Utilities (RPU) Data

We apply our model to data from a μPMU located near the Mountain View substation in Riverside, CA, from July 15 to August 15, 2017. The sensor reports four fundamental measurements on three phases: voltage magnitude, voltage phase angle, current magnitude, and current phase angle. The sampling rate is 120 Hz, but we aggregate the observations to a 1-second resolution to keep the computation feasible. This dataset has been analyzed before for understanding bank switching events (Shahsavari et al., 2017).

3.2. Experimental setup

Data preprocessing. Previous to performing the model inference, we preprocess the time series data to remove dominant trends and periodicity. In particular, we filter out the 100 frequencies with highest magnitude on the power spectral density of the time series. The purpose of these steps is to avoid discovering trivial fingerprints that are merely due to expected daily or weekly variations in energy usage. In Figure 3.2, we show the time series before (top) and after (bottom) preprocessing. We note that the the obvious periodic patterns have been dampened down while the large irregularities, such as the voltage drop in the beginning of August, are preserved.

Parameter setting. The two main parameters in our model are the two dispersion parameters α and λ, which we set to low values of 10^{-8} and 10^{-6}, respectively, with the goal to discover few dominant (and potentially interpretable) fingerprints which best characterize the data. For the prior distributions on μ and σ^2, we choose relatively "weak" hyperparameters, which are quickly superseded by the observed data.

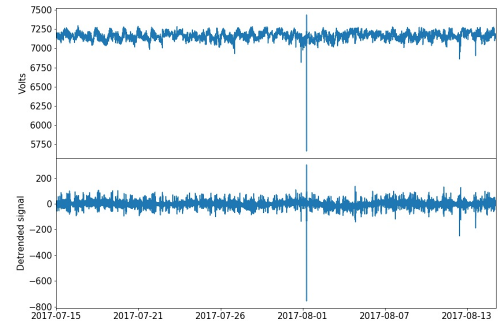

Figure 2. Top: Voltage time series from RPU. Bottom: We detrend the signal and filter out high-magnitude frequencies to remove periodicity.

3.3. Fingerprints capture different behaviors

In Figure 3.3, we show the maximum a-posterori (MAP) segments and fingerprints discovered by our model. Each color corresponds to a different fingerprint. We note that most of the time series is represented by a single fingerprint—in light blue. Segments generated by this fingerprint last for long periods of time, between one and two weeks. We interpret this fingerprint as a "normal" operating state of the transformer. Also of note is the red fingerprint, which starts at the beginning of August. There were multiple known disturbances in the RPU power grid at that time due to a lightning storm. Both a significant voltage drop on August 1 and smaller precursor events were observed. The green fingerprint is an event recurring at the same time step several days, and over a longer period on July 21-23. One interpretation for this pattern is a local control behind the point of measurement, such as a building management system turning on the air-conditioning. While this daily event is not of interest for reliability, it is an indicator of the health of the nearby system. For example if this signal were to change, i.e., the voltage drop increased over time, it would indicate an electromechanical fault.

We can also use the model to examine the degree of certainty of the segmentation by sampling from the posterior distribution of the hierarchical Dirichlet Process. In Figure 3.3, we show ten random samples of the MCMC sampler. We note that the beginning of August is always distinct from the surrounding times. Also, the two large segments around it are usually from the same fingerprint, but they are assigned to different fingerprints in some samples, such as samples 3 and 5 counting from the top. This suggests less certainty and the possibility of different activity patterns before and after the event in August.

4. Conclusions

We propose a Bayesian non-parametric model for unsupervised temporal signature discovery on μPMU data. Our

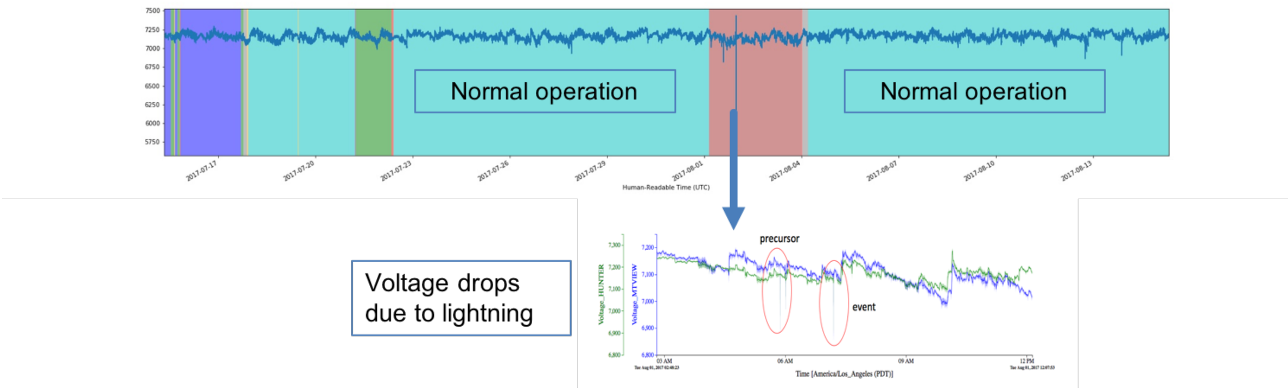

Figure 3. MAP estimator of the segments and fingerprints in the RPU data. Each color corresponds to a different fingerprint. The beginning of August is different from the surrounding times due to large voltage drops. The light blue fingerprint, which makes most of the time series, can be interpreted as a "normal" state.

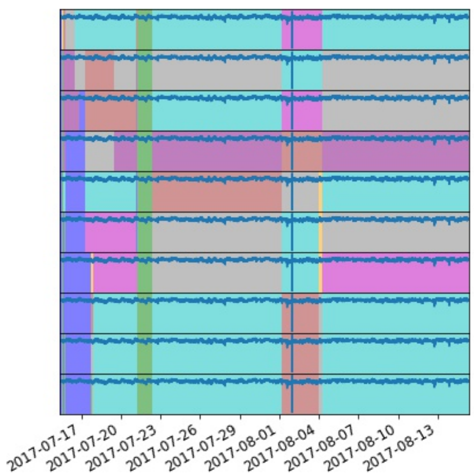

Figure 4. Samples from the posterior distribution. The beginning of August is consistently singled out as a different fingerprint.

model can identify different patterns of activity, such as normal operation and large voltage drops. In application, fingerprint discovery would be integrated in an alert system to monitor the "health" status of transformers in a power grid and take corrective actions as needed.

Limitations and future work. We assume that observations are independent draws given the fingerprint assignment, thus ignoring the temporal structure of the data. One promising extension is to model contiguous segments as samples of a Gaussian process (GP) with varying kernel parameters for each fingerprint. However, MCMC inference then becomes computationally expensive due to the matrix inversions needed to obtain the GP likelihood. We plan to use ideas from Bayesian change-point detection to perform inference without sampling.

References

Ali, S. et al. A machine learning approach to meter placement for power quality estimation in smart grid. *IEEE Transactions on Smart Grid*, 7(3):1552–1561, 2016.

Blei, D. M. and Frazier, P. I. Distance dependent chinese restaurant processes. *Journal of Machine Learning Research*, 12(Aug):2461–2488, 2011.

Diamantoulakis, P. D., Kapinas, V. M., and Karagiannidis, G. K. Big data analytics for dynamic energy management in smart grids. *Big Data Research*, 2(3):94–101, 2015.

Shahsavari, A. et al. A data-driven analysis of capacitor bank operation at a distribution feeder using micro-pmu data. In *2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–5. IEEE, 2017.

Shyam, R. et al. Apache spark a big data analytics platform for smart grid. *Procedia Technology*, 21:171–178, 2015.

Stewart, E. et al. Using micro-synchrophasor data for advanced distribution grid planning and operations analysis. Technical report, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2014.

Stewart, E. et al. Data-driven approach for monitoring, protection, and control of distribution system assets using micro-pmu technology. *CIRED-Open Access Proceedings Journal*, 2017(1):1011–1014, 2017.

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2006.

Zhang, H.-T., Xu, F.-Y., and Zhou, L. Artificial neural network for load forecasting in smart grid. In *2010 International Conference on Machine Learning and Cybernetics*, volume 6, pp. 3200–3205. IEEE, 2010.