Feature Extraction From Large-scale Data in Power Grid to Identify and Locate Events in Real-time

Wenting Li

Mentor: Deepjyoti Deka, Michael Chertkov

Rensselaer Polytechnic Institute (RPI), Troy, NY

Center for Nonlinear Studies (CNLS)-T4

August 9, 2018

Outline

1. Background
2. Low-dimensionality Structure of PMU Data & Fast Event Identification
3. Real-time Fault Location through Deep Learning
4. Conclusions and Future Work

Background

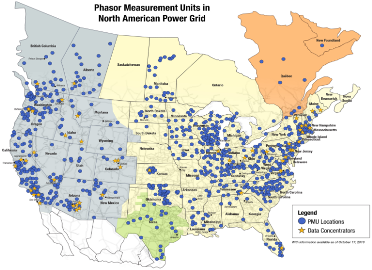


Figure: Installation of PMUs in the North America `https://www.naspi.org/documents`

- **Phase measurement units (PMU)** provide synchronized phasor measurements at the sampling rate of 30 or 60 sample per second;
- More than **2000** PMU are installed in the North America.
- PMU are generating large-scale of datasets in the power grid;
- **Data-driven** methods based on PMU data are promising to **automatically locate and identify** abnormal conditions in power grid.

Challenges and Opportunities

- For the **large-scale high dimensional** PMU data, how to extract useful information ?
- How to **connect the variations of data with the physical modeling** to monitor, adjust and optimize the current modeling ?
- How to propose **efficient and accurate algorithms** based on the PMU data to participate the closed-loop control?
- How to **reveal the correlations** in the data to augment our understanding of the system states?

Low-dimensional Structure of PMU data

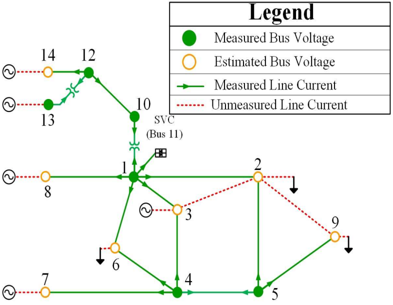


Figure: PMUs in Central NY Power Systems

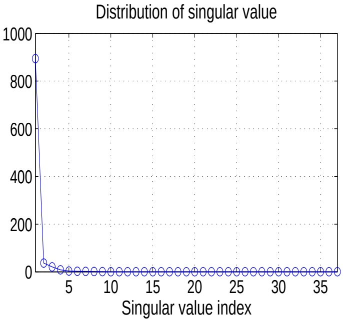


Figure: Singular values of the PMU data matrix

- 37 voltage/current phasors. 30 samples/second for 20 seconds. Form a large matrix of 600 by 37.
- Singular values decay significantly. Mostly close to zero. Singular values can be approximated by a sparse vector.
- Low-dimensionality also used in [Chen, Xie, Kumar 2013, Dahal, King, Madani 2012] for dimensionality reduction.

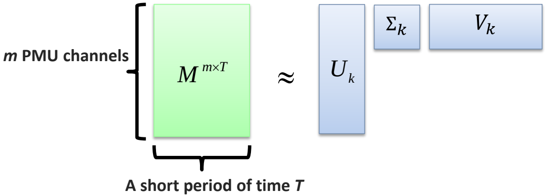
Motivations of Identifying Events

- Fast event identification is important to prevent cascading failures and enhance system security ;
- The existing data-driven methods have some limitations:
  - large dictionary size & complicated training: 6000 root patterns in a 21-bus system [Wang W et al., 2014], several hidden Markov models are trained to detect and identify events [Jiang H et al., 2014].
  - off-line algorithms & long window size: [Song Y et al., 2015] employed an off-line algorithm of 20-second data ;
  - high sampling frequency: [Jiang H et al., 2014] utilized 1kHz sampling rate based on frequency disturbance recorder;

Problem Formulation and Feature Extraction

- **Problem**: Given the PMU data $M \in R^{m \times T}$ of m buses in the time period T, we want to identify events like line trip, generator trip, line fault.
- **Main Idea**: Extract features and establish a dictionary;
- Extract **feature matrix** $V_k \in R^{T \times k}, k \ll m, k \ll T$ by singular value decomposition (SVD) in (1);

$$M = U_k \Sigma_k V_k \tag{1}$$



U_k, V_k : span column and row subspaces;
Σ_k : the k largest singular values of M.

Physical interpretation of feature matrix V_k

- A linear model after an impulse input follows

$$
\begin{aligned}
x(t+1) &= Ax(t) + Bu(t) \\
y(t) &= Cx(t)
\end{aligned}
$$

  where $x(t), y(t)$ are the deviation of state variable and measurements at time t, $u(t)$ is the input variable, and A, B, C are the state matrix, input matrix and observation matrix
- Let $\boldsymbol{\beta}_k^{\dagger} = [\lambda_k, \lambda_k^2, \cdots, \lambda_k^T]$, where λ_k are the kth eigenvalue of A;
- Span(V_k) = Span($\boldsymbol{\beta}$) assume that in a short time $u(t) \sim 0$
- Various types of events excite distinctive eigenvalues thus V_k are different;
- V_k can represent different dynamics after events.

Our Approach: Event Identification through Dictionary

- Identify an event by comparing the row subspace of the real-time spatial-temporal PMU data blocks with a dictionary of subspaces obtained from recorded PMU data with known event types.

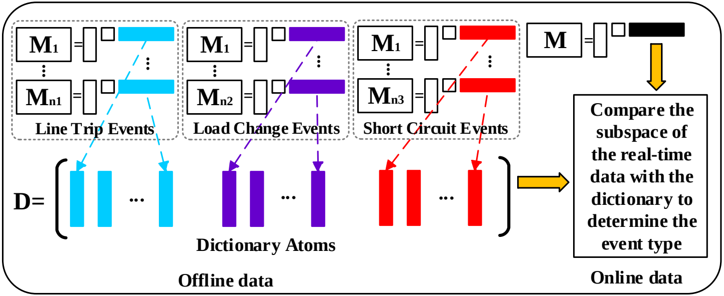


Figure: Dictionary construction from historical datasets and real-time data identification through subspace comparison

Li W, Wang M, Chow JH. Fast event identification through subspace characterization of pmu data in power systems. IEEE PES General Meeting, Chicage,IL, July 2017 (pp. 1-5).

Li W, Wang M, Chow JH. Real-time Event Identification through Low-dimensional Subspace Characterization of High-dimensional Synchrophasor Data. IEEE Transactions on Power Systems. 2018 Jan 22.

Similarity of Subspaces

- Subspace Angle [Mahdi S et al., 2014]

$$\theta(\mathcal{S}_l, \mathcal{S}_k) = \arccos(\sqrt{\|B_k^\dagger B_l\|_F^2 / \min\{k, l\}}). \tag{2}$$

- θ equals 90° if two subspaces are orthogonal to each other;
- θ equals 0° if two subspaces are the same $(l = k)$ or one is embedded in the other $(l \neq k)$.
- A smaller θ indicates higher affinity of two subspaces.

Experimental Results of ISO NE PMU data

Table: Minimum subspace angles between a test case and the dictionary atoms of three types of events in recorded PMU data

Events \ Dictionary	Load Change	Fault	Line Trip
Load Change 1	$\mathbf{4.08}^\circ$	16.91°	18.29°
Load Change 2	$\mathbf{3.12}^\circ$	20.81°	14.39°
Fault 1	24.95°	$\mathbf{6.33}^\circ$	23.86°
Fault 2	8.93°	$\mathbf{3.73}^\circ$	15.76°
Line Trip 1	7.25°	5.85°	$\mathbf{3.93}^\circ$
Line Trip 2	11.20°	30.21°	$\mathbf{4.27}^\circ$

- The minimum subspace angle (bolded) indicates the type of the events;
- 32 events of three types are tested with 100 % identification accuracy rate.

Motivations of Locating Faults

- Locating faults in real time is crucial to improve the power system **stability and reliability**;
- Impedance-based methods often **assume loads are static** and are sensitive to topology change;
- Traveling-wave-based method require **high sampling rate and accuracy** of measurements;
- Artificial intelligent methods: some require high sampling rate like 2400 Hz (Mehrdad 17), some are DC model based, some only **for single type** of faults or single transmission line and **complete observability of the system** required (Guangyu 16).

Problem Formulation

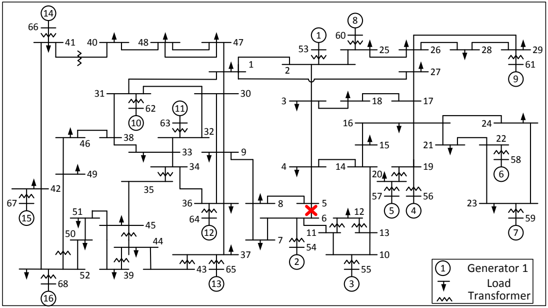


Figure: The line 5-6 is faulted in the IEEE 68-bus power system

- **Problem:** When a line is faulted (marked as red cross), how to locate the fault efficiently?
- **Challenges:**
  - The **type** of fault can be various, including symmetrical or asymmetrical fault;
  - Various **fault impedances** cause voltage drop in different degree;
- **Our Approach:**
  - Extract location features;
  - Classify by a convolutional neural network (CNN).

Feature Extraction

Given voltage PMU data of the power system with n buses before and during fault $U^0, U' \in \mathcal{C}^n$ and admittance matrix $Y^0 \in \mathcal{C}^{n\times n}$ before the fault and $\Delta U = U^0 - U'$, we define the **feature vector** $\psi \in \mathcal{C}^n$ in (3):

$$\psi = Y^0 \Delta U \tag{3}$$

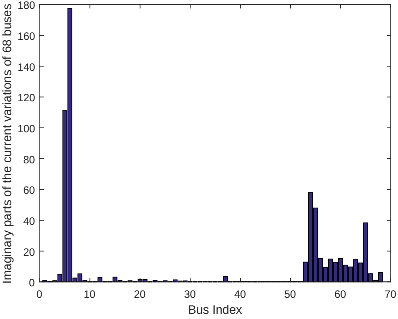


Figure: The imaginary parts of the feature vector ψ after the **line 5-6** is faulted

- **Physical Interpretation of ψ:** Based on the **substitution theory**,

$$\psi = \Delta I^u + \Delta I \tag{4}$$

where ΔI^u is a sparse vector with nonzero values only at the terminal buses of the faulted line;
ΔI denotes the current variations of buses.

Our CNN Classifier

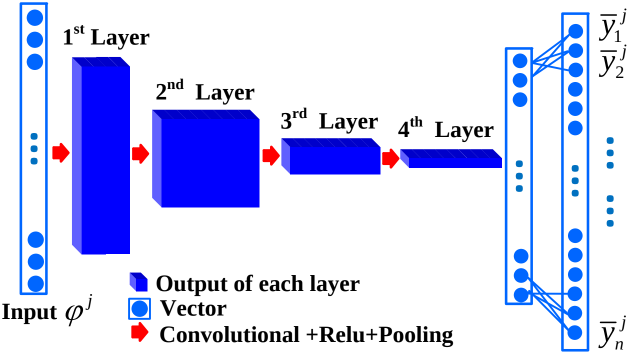


Figure: Understanding CNN

- Input the extracted feature ψ^j and the label y^j of the jth dataset;
- CNN optimizes the parameters by minimizing a loss function, and then outputs $\bar{y}_i^j$, **the probability of the ith line** for the jth dataset;
- The line with the highest probability indicates the faulted line.

Numerical Results

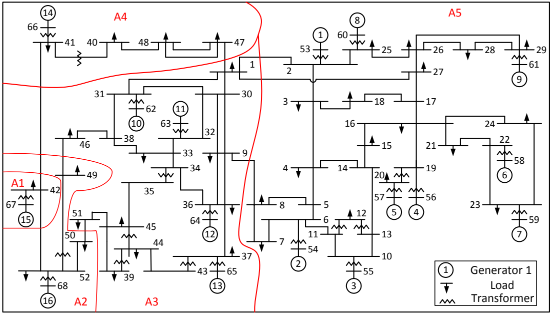


Figure: The IEEE 68-bus Power System

- Four types of faults, including three phase (**TP**), line to ground (**LG**), double line to ground (**DLG**) and line to line (**LL**) faults, are simulated in the IEEE 68-bus power system;
- More than 2300 datasets of **various locations, different types and fault impedances, and random load fluctuations** are generated;
- Data rate is 30 samples per second.

Location Accuracy with Complete Measurements

Table: The LAR η ($\%$) of **MSVM**

Z_f (p.u.)	0.05	0.01	0.001	0.0001
LAR of TP ($\%$)	100	100	100	100
LAR of LG ($\%$)	100	100	100	100
LAR of DLG ($\%$)	98.6	100	100	99.5
LAR of LL ($\%$)	98.6	99.6	93.5	94.6

Table: The LAR η ($\%$) of **CNN or NN**

Z_f (p.u.)	0.05	0.01	0.001	0.0001
LAR of TP ($\%$)	100	100	100	100
LAR of LG ($\%$)	100	100	100	100
LAR of DLG ($\%$)	100	100	100	100
LAR of LL ($\%$)	100	100	100	100

- The location accuracy rate (**LAR**)[1] of Multi-class support vector machine (**MSVM**), Neural Network(**NN**) and **CNN** are all more than 90% for most events when the system is **completely measured.**

[1] defined as $\eta = \frac{\text{The number of faults correctly located}}{\text{total number of faults}}$

Location Accuracy with Partial Measurements

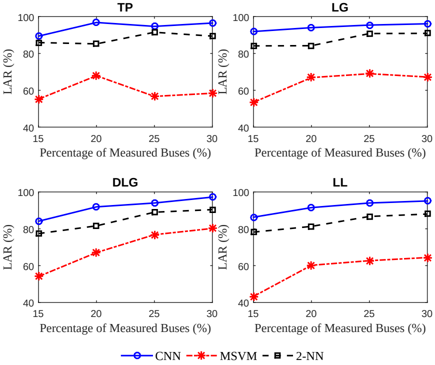


Figure: The LAR of the CNN, MSVM, NN with different percentage of measured buses

- When **15** $\%\sim$ **30** $\%$ buses are measured, the LAR of CNN is **higher** than that of MSVM and NN for different types of faults;
- When at least **30** $\%$ buses are measured, the location accuracy of CNN can be higher than 95%;
- What if less than **15** $\%$?

Performance When $\leq 15\%$ Buses are Measured

Table: The ARC of CNN on different types of events with the ratio of measured buses less than 15%

Measured Ratio	TP	LG	DLG	LL
7 $\%$	1.32	1.48	1.92	1.56
10 $\%$	1.38	1.28	1.66	1.54
15 $\%$	1.38	1.23	1.57	1.54

- Define Averaged Rank of the Correct (**ARC**) line to indicate how many lines that mostly include the correct lines;
- **The ARC, less than 3**, when $\leq 15\%$ ratios of buses are measured, indicates that most correct lines are in the lines having the **top-3 probability**;
- **Moreover**, the lines with high probability are mostly near the faulted line.

Example: Line 5-6 is faulted

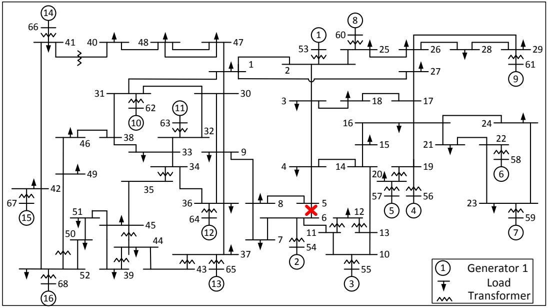


Figure: The lines with the top-5 probability when only 5 buses are measured

- Only **7**% buses are measured;
- The correct line 5-6 has the **second highest probability**;

The Lines with Top-5 Probability

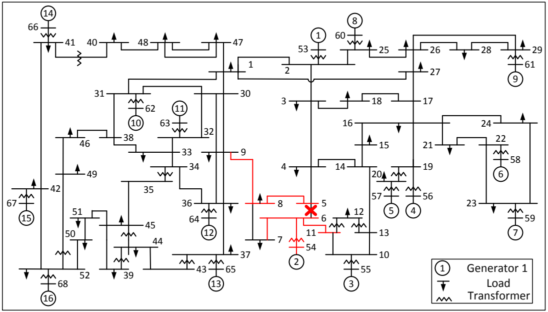


Figure: The lines with the top-5 probability when only 5 buses are measured

- The lines with the top-5 probability marked in red are **in the neighborhood** of the faulted line 5-6;
- This neighborhood property is **not a special case** but commonly exist in most cases;
- **Why** the lines with high probability show the neighborhood property? is it coincidence or due to some magic of CNN?

The essensity of the neighborhood property

- It is verified that the neighborhood property exist even **other classifier like NN** is applied;

The essensity of the neighborhood property

- It is verified that the neighborhood property exist even **other classifier like NN** is applied;
- The essential cause is due to the features we defined:

$$\psi = \Delta I^u + \Delta I \tag{5}$$

The essensity of the neighborhood property

- It is verified that the neighborhood property exist even **other classifier like NN** is applied;
- The essential cause is due to the features we defined:

$$\psi = \Delta I^u + \Delta I \tag{5}$$

- For the kth ($k \neq i, j$) entry ψ_k when the line i-j is faulted:

$$\psi_k = \Delta I_k + \Delta I_k^u = \Delta I_k \tag{6}$$
$$= \Sigma_{j \in \mathcal{N}_k} Y_{kj}^0 \Delta U_j \tag{7}$$

where Y_{kj}^0 denotes the admittance between bus k and j before the fault, and $\mathcal{N}_k$ denotes **the neighbor of the bus** k.

The essensity of the neighborhood property

- It is verified that the neighborhood property exist even **other classifier like NN** is applied;
- The essential cause is due to the features we defined:

$$\psi = \Delta I^u + \Delta I \tag{5}$$

- For the kth ($k \neq i, j$) entry ψ_k when the line i-j is faulted:

$$\psi_k = \Delta I_k + \Delta I_k^u = \Delta I_k \tag{6}$$

$$= \Sigma_{j \in \mathcal{N}_k} Y_{kj}^0 \Delta U_j \tag{7}$$

  where Y_{kj}^0 denotes the admittance between bus k and j before the fault, and $\mathcal{N}_k$ denotes **the neighbor of the bus** k.
- Thus ψ_k denotes the sum of the total current variations in the neighborhood of k.

Greedy Algorithm of PMU Placement

Algorithm 1 Greedy Algorithm to Select PMU Placement

Input parameters: $m, \beta = 1$
Initialize : $\mathcal{S}_0 = \emptyset$ and the loss value $l = \infty$
for $k = 0, \cdots, m$ **do**
  for bus $i = 1, \cdots, n$ **do**
    Compute the minimum loss $l_i = \min_\Theta l(\Theta, \{\mathcal{S}_k \cup i\})$ of (1) for each bus i
  end for
  $i^* = \arg\min_i (\frac{\beta}{d_i} + l_i)$,
  where d_i the degree of bus i, β is a weight parameter.
  if $l_{i^*} < l$ **then**
    $\mathcal{S}_{k+1} = \mathcal{S}_k \cup i^*$
    $l = l_{i^*}$
  else
    $\mathcal{S}_{k+1} = \mathcal{S}_k$
  end if
end for
Output: $\mathcal{S}_m$

The loss function is (1) is:

$$l'(\Theta, \mathcal{S}) = \min_\Theta \frac{1}{N} \Sigma_{j=1}^{N} \Sigma_{i=1}^{n} y_i^j \log \bar{y}_i^j (\Theta, \mathcal{S}) + \lambda \|\Theta\|_F^2$$

Performances Comparison of Different ALgorithms

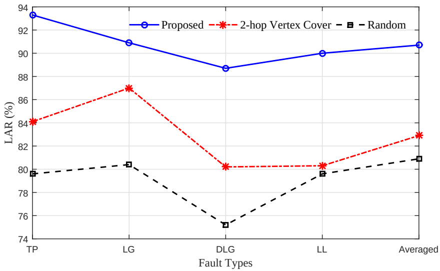


Figure: LAR of CNN based on different algorithms

- The LAR of the four types of faults are compared when about 15% buses are measured by different algorithms;
- The "2-hop Vertex Cover" method is only based on the graph connectivity and "Random" method is to select buses arbitrarily;
- The proposed algorithm achieves a better location accuracy rate for different types of faults.

Conclusions & Future Work

- **A real time** fault location method is proposed based on a CNN classifier;
- This method can keep a high location accuracy when the system is **fully or partially** measured;
- Our CNN classifier achieves a higher LAR than that of the MSVM and NN when the system is partially measured;
- **The lines with high probabilities** can indicate the small area of the faulted line;
- The proposed PMU placement algorithm can improve the location accuracy rate;
- More sensitivity of our method will be tested in the **future work**.

References

Majidi, Mehrdad and Etezadi-Amoli, Mehdi and Fadali, Mohammed Sami (2017)
A sparse-data-driven approach for fault location in transmission networks
IEEE Transactions on Smart Grid , 548–556.

Garcia, Manuel and Catanach, Thomas and Vander Wiel, Scott and Bent, Russell and Lawrence, Earl (2016)
Line outage localization using phasor measurement data in transient state
IEEE Transactions on Power Systems, 3019–3027.

Feng, Guangyu and Abur, Ali (2016)
Fault location using wide-area measurements and sparse estimation
IEEE Transactions on Power Systems, 2938–2945.

Koley, Ebha and Yadav, Anamika and Thoke, Aniruddha Santosh (2015)
A new single-ended artificial neural network-based protection scheme for shunt faults in six-phase transmission line
International Transactions on Electrical Energy Systems, 1257–1280.

References

Wei Wang, Li He, Penn Markham, Hairong Qi, Yilu Liu, Qing Charles Cao, and Leon M Tolbert (2014)

Multiple event detection and recognition through sparse unmixing for high-resolution situational awareness in power grid

IEEE Transactions on Smart Grid, 5(4):16541664, 2014

Mahdi Soltanolkotabi, Ehsan Elhamifar, Emmanuel J Candes, et al (2014)

Robust subspace clustering

The Annals of Statistics,, 42(2):669699, 2014.

Yang Song, Wei Wang, Zhifei Zhang, Hairong Qi, and Yilu Liu (2015)

Multiple event analysis for largescale power systems through cluster-based sparse coding

IEEE International Conference on Smart Grid Communications , 301306. IEEE, 2015.

Huaiguang Jiang, Jun Jason Zhang, Wenzhong Gao, and Ziping Wu. (2014)

Fault detection, identification, and location in smart grid based on data-driven computational methods

IEEE Transactions on Smart Grid, 13, 5(6):29472956, 2014.

Acknowledge

- Deepjyoti Deka, Michael Chertkov, Garcia, Manuel
- CNLS and LANL