FOA 1861 FINAL PROJECT BRIEFING
BIG DATA ANALYSIS OF SYNCHROPHASOR DATA

A Robust Event Diagnostics Platform: Integrating Tensor Analytics and Machine Learning into Real-time Grid Monitoring

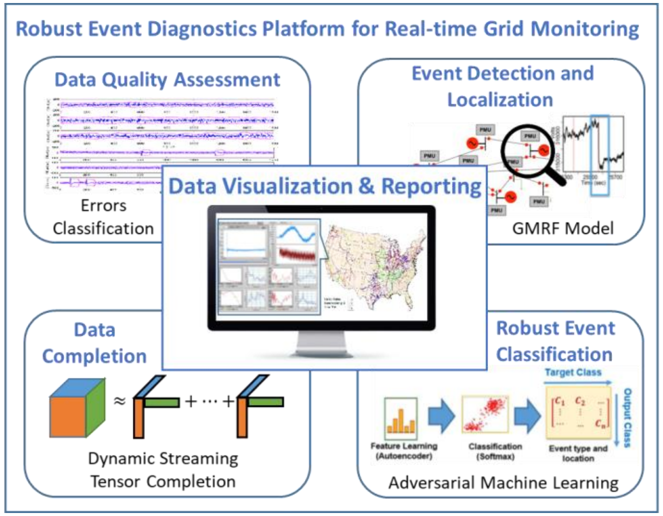
Lei Yang
University of Nevada, Reno
leiy@unr.edu

July 28, 2021

Project Partners
Hanif Livani (UNR), Virgilio Centeno (VT), Junshan Zhang (ASU), Pin-Yu Chen (IBM)

Project Overview

- **Project Objectives:**
  - Data quality assessment will flag and recover bad data using tensor completion techniques.
  - Classify event signatures and detect system wide events.



Overview of key functionalities of the robust event diagnostics platform.

- **Technical Approach:**
  - Tensor analytics-based PMU data completion approach by utilizing both spatial and temporal correlation of PMU data.
  - Machine learning based robust event classification.

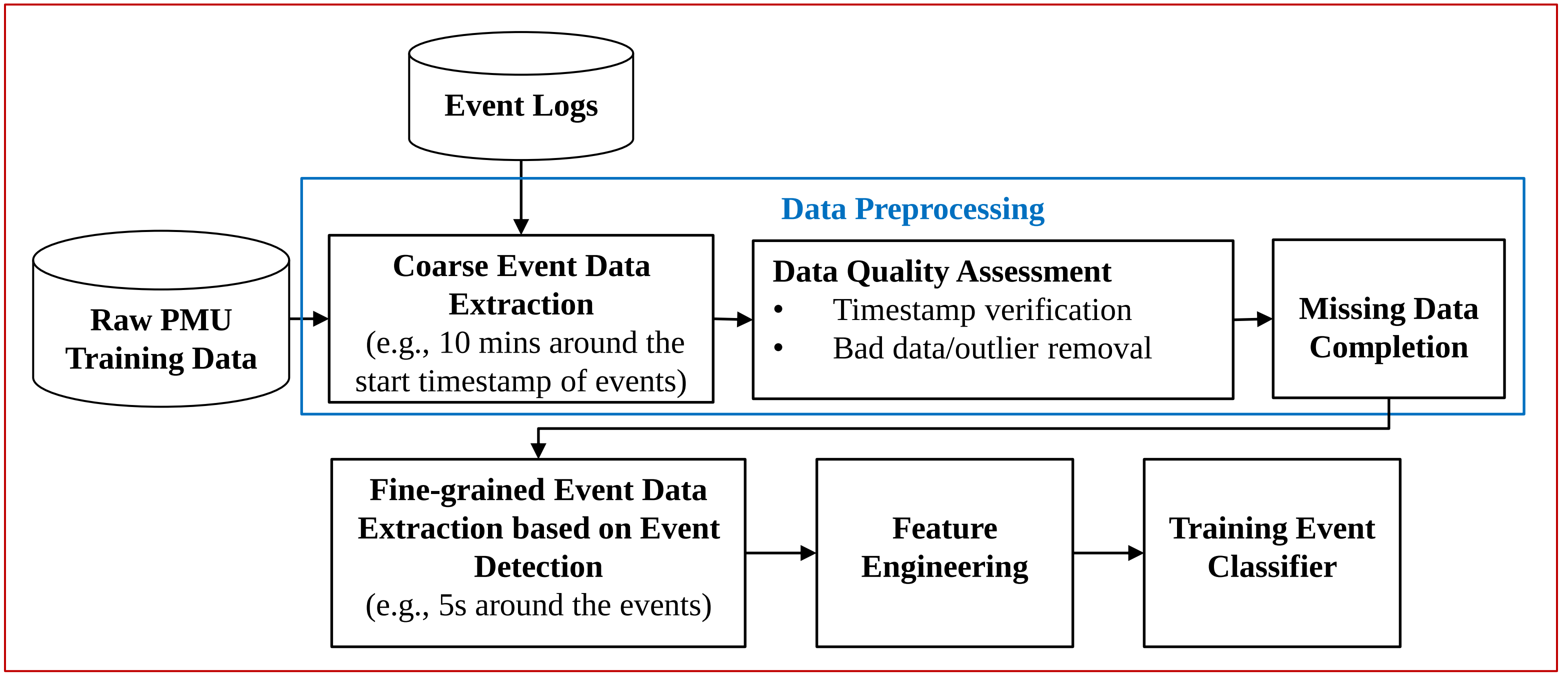
- **Significance & Impact:**
  - By integrating *Tensor Analytics and Machine Learning*, this project will provide innovative tools for PMU data management, which can assist grid operators to better assess the state of the system and contribute to the efficient, safe, reliable operation and design of the electric system.

Technical Accomplishments

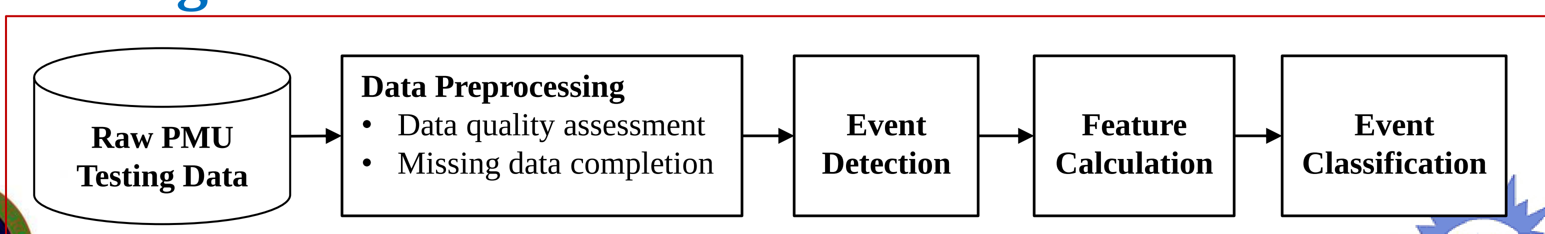
- **Specific goals/objectives reached**
  - **Data Quality Assessment and Completion**
    - We performed data quality assessment on the PMU datasets.
    - We developed an algorithm for unwrapping angles in presence of missing data.
    - We developed the regularized tensor completion method.
  - **Robust Event Diagnostics**
    - We developed an event detection method based on low rank property of PMU data and did the event detection experiments for the entire IC-B dataset.
    - We performed comprehensive data exploration and preprocessing on PMU datasets.
    - We developed machine learning based event classification models.

Outline of Technical Approaches

Training

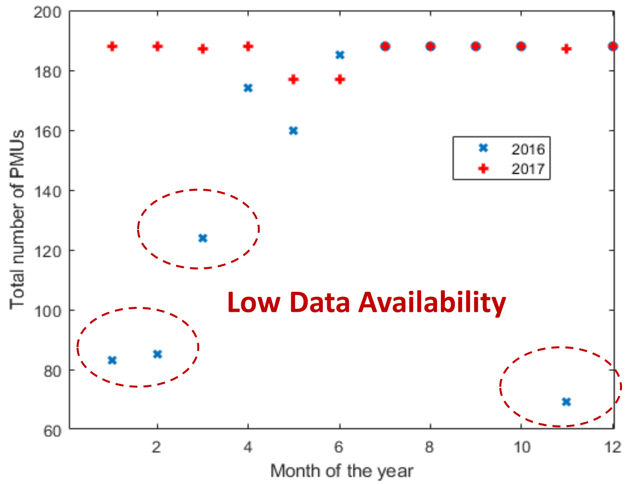


Testing

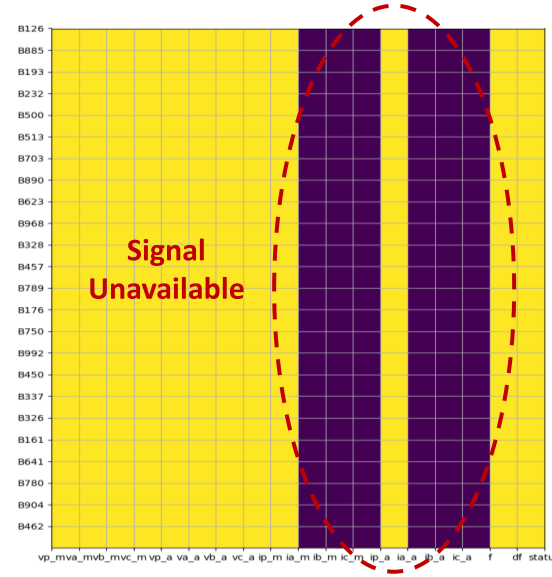


Experimental Results
--Data Quality Assessment (1)

- Not all PMUs report data for every day and Not all PMUs report all measurements.

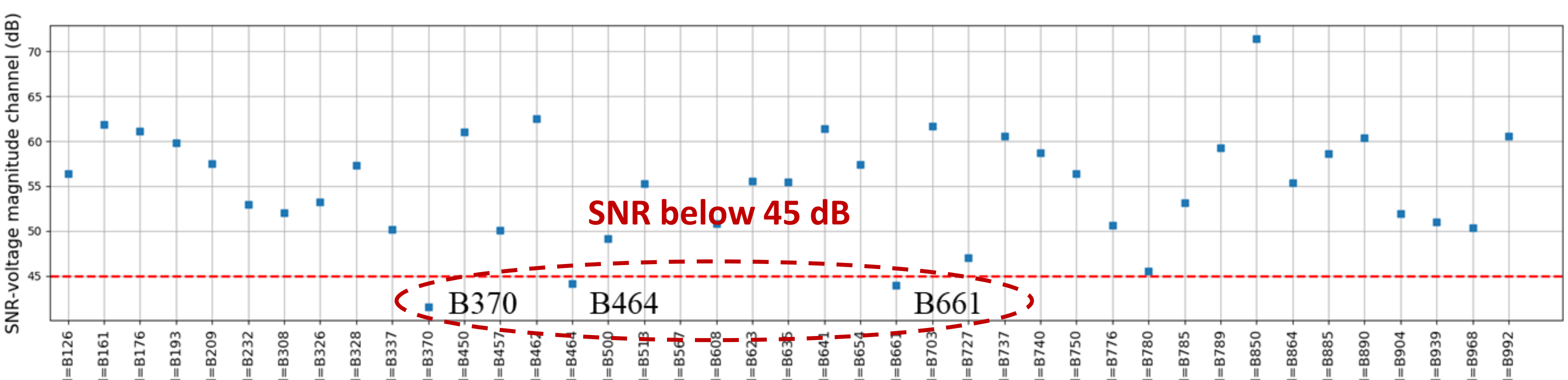


PMU availability for IC-C



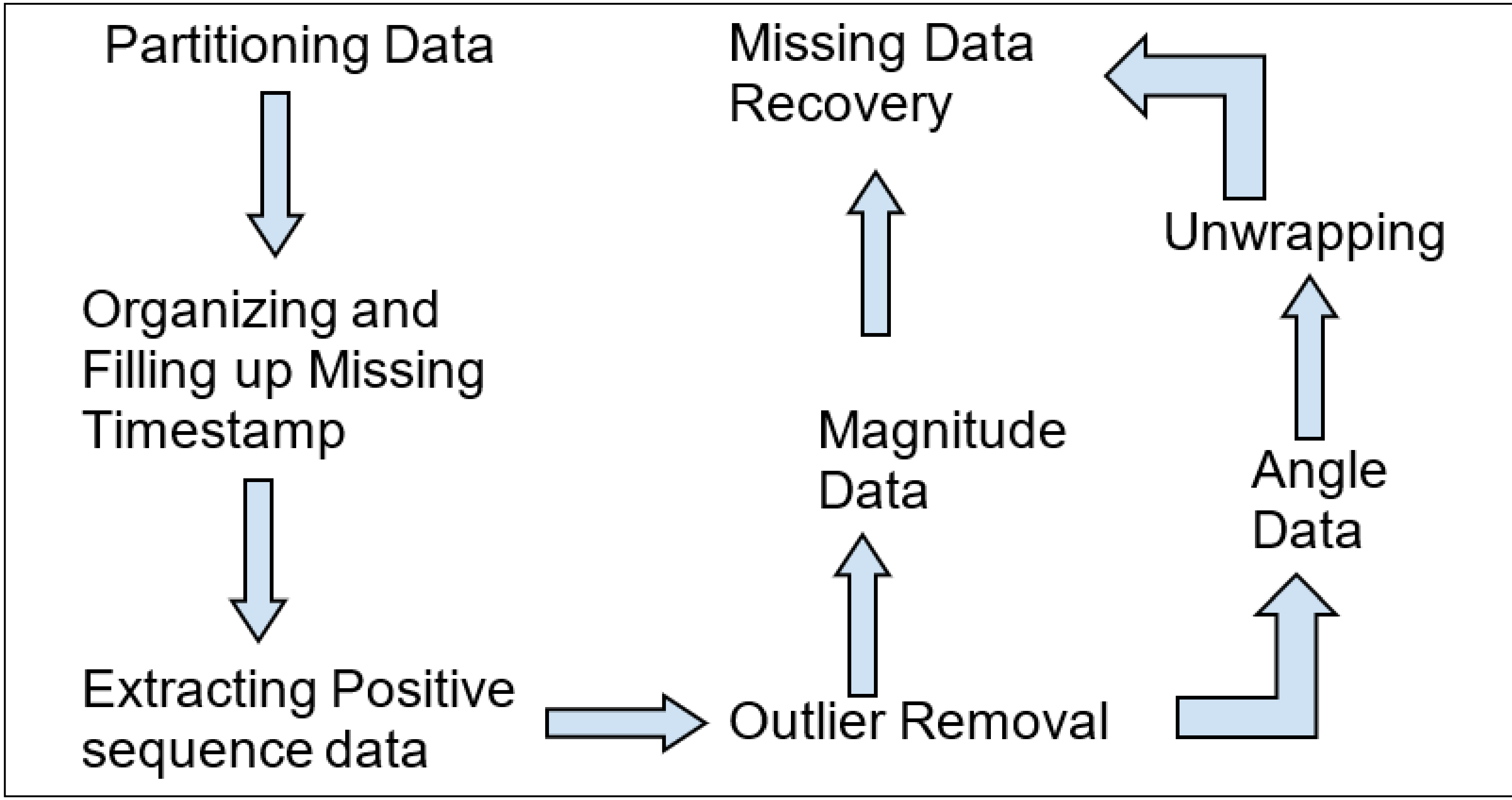
Signal availability in 60 fps PMUs for IC-B

- Low SNR for some PMUs



Experimental Results
--Data Quality Assessment (2)

- Data Cleaning Process



Experimental Results --Missing Data Completion (1)

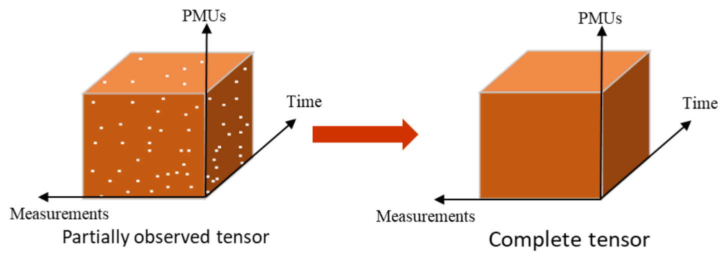
- Regularized Low Rank Tensor Completion (LRTC) for PMU Data

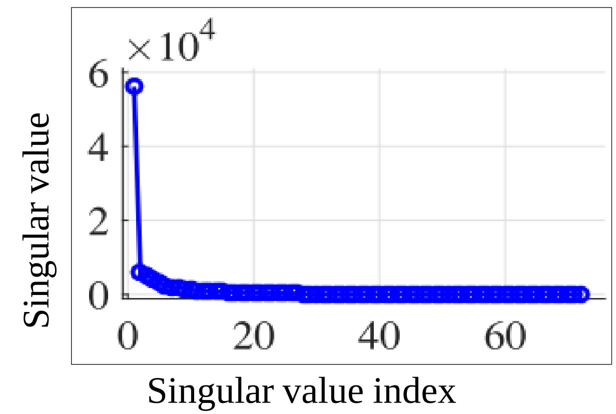
Data dependency matrices

Regularized LRTC

$$
\begin{cases}
\min_{\mathcal{X}, U_k} rank(\mathcal{X}) \\
s.t. \quad \mathcal{M}_\Omega = \mathcal{X}_\Omega \\
\mathcal{X} \approx [\![U_1, U_2, U_3]\!] \\
\left\| \left(X_{(k)} A_k \right) - L_k \right\|^2 = 0
\end{cases}
$$

Matrices representing natural constraints





Low rank property of PMU measurements

- Validation Results
  - Missing randomly
  - Multiple channel missing

Experimental Results
--Missing Data Completion (2)

- **Validation Results:**
  - The measurements are from interconnect B, and 23 PMUs during a line outage fault for 5s with sampling rate of 60 samples per second.

Missing randomly

	Methods	$p = 0.1$	$p = 0.2$	$p = 0.3$	$p = 0.5$
Tensor-based Methods	CP-ALS	0.0010	0.0015	0.0025	0.0035
	TNCP	0.0010	0.0016	0.0031	0.0079
	Reg. LRTC	**0.0006**	**0.0014**	**0.0018**	**0.0024**
Matrix-based Methods	Meanfill	0.313	0.444	0.541	0.706
	NNM	0.0062	0.077	0.189	0.804
	MMMF	0.0090	0.012	0.016	0.022
	KNN	0.0019	0.0044	0.0063	0.0129

Multiple channel missing

	Methods	1 CM	3 CM	6 CM	12 CM
Tensor-based Methods	CP-ALS	0.0001	0.0004	0.0165	0.0501
	TNCP	0.0002	0.0103	0.0123	0.0605
	Reg. LRTC	**0.0001**	**0.0004**	**0.0015**	**0.0043**
Matrix-based Methods	Meanfill	0.0914	0.1686	0.2696	0.3795
	NNM	0.0366	0.0841	0.1597	0.2533
	MMMF	0.0052	0.0091	0.0134	0.0218
	KNN	0.0006	0.0007	0.0019	0.0073

- Our method outperforms all the other methods in terms of accuracy.
- Tensor based methods are capable of recovering the missing values for consecutive missing entries.

A. Ghasemkhani, I. Niazazari, Y. Liu, H. Livani, V. A. Centeno and L. Yang, "A Regularized Tensor Completion Approach for PMU Data Recovery," in IEEE Transactions on Smart Grid, doi: 10.1109/TSG.2020.3030566.

Experimental Results --Event Classification (1)

Fine-grained Event Data Extraction based on Event Detection (e.g., 5s around the events) → **Feature Engineering** → **Training Event Classifier**

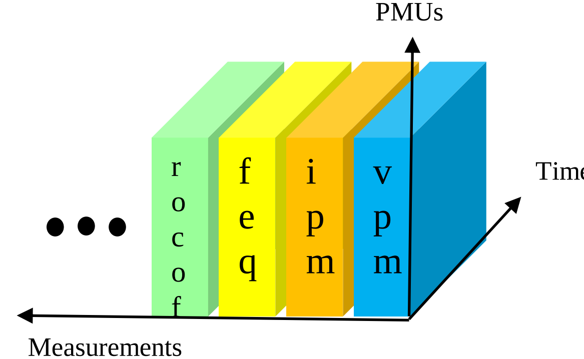
- **Event detection based on the change of singular values**
  - Ratio of the largest σ_1 and second largest σ_2 singular values of PMU matrices

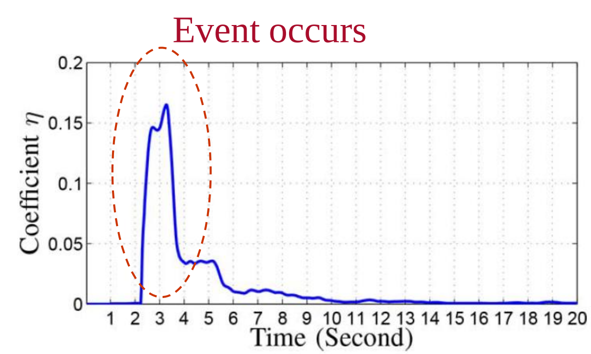
$$\boxed{\eta_t = \frac{\sigma_2}{\sigma_1}}$$

  - Change of this ratio from time $\boldsymbol{t}$ to time $\boldsymbol{k}$

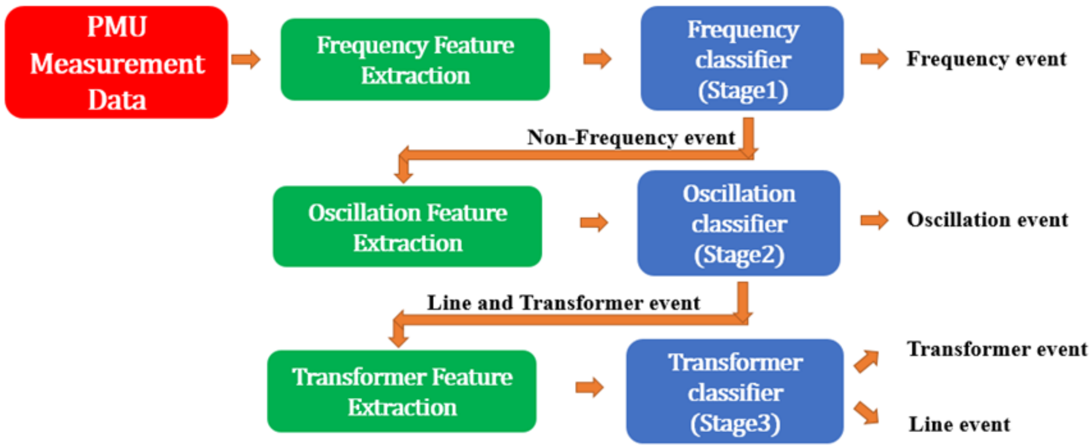
$$\boxed{\xi_{t,k} = \frac{\eta_t - \eta_k}{\eta_k (t-k)}}$$

  - Events are detected if $\xi_{t,k}$> threshold.
  - Bayesian optimization-based threshold tuning
- **98% detection accuracy on IC-B dataset with 4854 events**

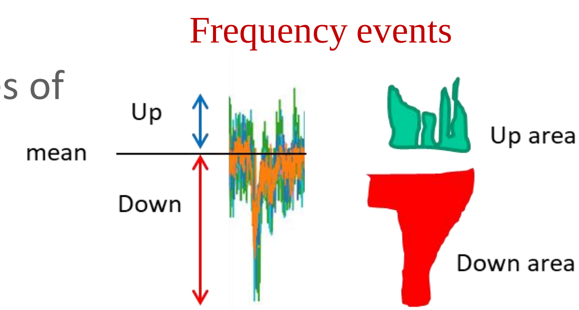




Experimental Results
--Event Classification (2)



- **Event classification**
  - **Feature Engineering**: construct features for different types of events.
  - **Hybrid event classification model.**
  - **Validation results**:
    - **98% accuracy for frequency events**
    - **91% accuracy for all events**, compared to 83% using CNN trained with original PMU measurements.



- **Improve interpretability of data-driven models for power system event classification.**

Value of Work

- **Benefits for the utilities**
  - Utilities can use our methods to enhance the quality of the collected PMU data, detect and classify events in real time, which can assist grid operators to better assess the state of the system and contribute to the efficient, safe, reliable operation and design of the electric system.
  - High-value use cases for the tools
    - Missing data completion
    - Event detection
    - Event classification

- **Benefits of sharing data from multiple utilities**
  - **Establish data collection standards**
    - How to store the PMU data to facilitate ML and BD analytics
    - How to prepare the event logs: The event logs from multiple utilities are of different quality in terms of event descriptions.

- **It is worthwhile for utilities to share anonymized data.**
  - The limitation of anonymized data is that it may limit the performance of the developed methods (e.g., event classification).

Readiness for Commercialization

- **Readiness number of the product developed in this project** (0 being fundamental early-stage research and 10 being commercially viable for sale as a working product)
  - **Missing data completion: 5**
  - **Event detection: 6**
  - **Event classification: 5**

- **Next steps**
  - Submit white papers for follow-up work with utilities to make the results from the projects available to use and transition our research to tools for utilities.

Being Ready for ML & BD Analytics (1)

- **Off-the-shelf machine learning models cannot achieve good performance for PMU data analytics.**

- **Key challenges of AI/ML methods and possible solutions**
  - **Challenge 1: Missing data (consecutive missing data)**
    - **Solution:** Regularized Tensor Completion for PMU Data by utilizing both spatial and temporal correlation of PMU data
  - **Challenge 2: Inaccurate timestamps of event logs**
    - **Solution:** Fine-grained Event Data Extraction based on Event Detection using low rank property of PMU data
  - **Challenge 3: What features to extract?**
    - **Solution**: Feature engineering: Different features for different events (e.g., ROCOF for frequency events)
  - **Challenge 4: Unbalanced training data**
    - For example, in IC-B, we have 4854 events, where the number of line outage events is about 75% in the event logs and the number of oscillation events is about 100. You will see overfitting when applying off-the-shelf machine learning models.
    - **Solution: Sampling methods and data augmentation**
  - **Challenge 5: Need additional information**
    - Event Information: The line events and transformer events are similar, which are difficult to classify. We may need to know more information for these events in order to separate these events.
    - Topology information
    - Information of PMU measurements, e.g., specific current phase sequence

Being Ready for ML & BD Analytics (2)

- **Recommendations for PMU data collection and event logs.**
  - Establish data collection standards and add more descriptions of events in the event logs, to help the development of ML models.
    - E.g., the oscillation modes of oscillation events in IC-B are not given.
  - Improve the timestamps of the events in the event logs.
  - Provide more information of the systems

- **It is important to synergistically combine machine learning models with power systems domain knowledge.**
  - Our methods combine machine learning models with domain knowledge to improve the performance.

- **Recommended low-cost steps for utilities to use big data analytics 2 or 3 years from now.**
  - Establish data collection standards and improve the quality of event logs.
  - Study the feasibility of applying the tools developed in FOA 1861 projects in real system.

Lessons Learned and Next Steps

- **Recommendations for using the FOA 1861 dataset**
  - Improve the event logs for future use
- **Distinguish significant threats from the common disturbances that happen every day**
  - Our detection algorithm outputs high values for significant threats, compared to low values for common disturbances.
  - Score based classification to distinguish significant threats from common disturbances.
- **Improve interpretability of data-driven models for power system event detection and classification**
  - Use power system’s domain knowledge to create features for event detection and classification
- **Next steps for ML & BD analytics**
  - Automated data labeling
  - Adversarial machine learning subject to data privacy and security

Publications

- A. Ghasemkhani, A. Darvishi, I. Niazazari, A. Darvishi, H. Livani, and L. Yang, “Deepgrid: Robust deep reinforcement learning-based contingency management,” in 2020 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT), February 2020.
- Iman Niazazari, Hanif Livani, “Attack on Grid Event Cause Analysis: An Adversarial Machine Learning Approach,” in 2020 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT), February 2020.
- Pu Zhao, Pin-Yu Chen, Siyue Wang, and Xue Lin, “Towards Query-Efficient Black-Box Adversary with Zeroth-Order Natural Gradient Descent,” AAAI, Feburary 2020.
- Cheng, Minhao, Simranjit Singh, Pin-Yu Chen, Sijia Liu, and Cho-Jui Hsieh. "Sign-OPT: A Query-Efficient Hard-label Adversarial Attack." *Accepted by ICLR* 2020.
- A. Ghasemkhani, I. Niazazari, Y. Liu, H. Livani, V. A. Centeno and L. Yang, "A Regularized Tensor Completion Approach for PMU Data Recovery," in IEEE Transactions on Smart Grid, doi: 10.1109/TSG.2020.3030566.
- Tsuyoshi Idé, Amit Dhurandhar, Jiri Navratil, Moninder Singh, Naoki Abe, “Anomaly Attribution with Likelihood Compensation,” In Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI 21, February 2-9, 2021, virtual), pp.TBD, to appear.
- I. Niazazari, H. Livani, A. Ghasemkhani, Y. Liu, and L. Yang, “Event cause analysis in distribution networks using synchro waveform measurements,” in 2020 North American Power Symposium (NAPS), April 11, 2021.
- I. Niazazari, Y. Liu, A. Ghasemkhani, S. Biswas, H. Livani, L. Yang, and V. A. Centeno, “PMU-data-driven event classification in power transmission grids,” in 2021 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT), February 2021.
- A. Ghasemkhani, Y. Liu, and L. Yang, “Low-rank tensor completion for PMU data recovery,” in 2021 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT), February 2021.

Thank you!

Question?

Lei Yang, leiy@unr.edu