FOA 1861 FINAL PROJECT BRIEFING
BIG DATA ANALYSIS OF SYNCHROPHASOR DATA

Robust Learning of Dynamic Interactions for Enhancing Power System Resilience

Yuxuan Yuan
Iowa State University
yuanyx@iastate.edu
April 12, 2022

Dr. Zhaoyu Wang, PI, ISU
Dr. Ian Dobson, Co-PI, ISU
Dr. Venkataramana Ajjarapu, Co-PI, ISU
Dr. Jie Chen, Co-PI, IBM
Neeraj Nayak, Co-PI, EPG
Joel Lindsay, NETL
Sandra Jenkins, DOE Project Officer

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

NETL
NATIONAL ENERGY TECHNOLOGY LABORATORY

Outline

- Project Overview
- Experimental Results
- Technical Accomplishments
- Value of Work
- Readiness for Commercialization
- Readiness for ML & BD Analytics
- Lessons Learned and Next Steps
- Publications List

Project Overview

The overall goal of the project is to leverage robust graphical learning and PMU data to learn the dynamic interactions of electrical grid components in order to improve the power system resilience. Specifically, this project incorporates four objectives:

1) Massive PMU data preparation, refining, and real-time visualization and access.
2) Identifying and cataloguing anomalous patterns.
3) Learning interaction graphs using deep graph neural networks.
4) Graph-based modeling, monitoring, and mitigation of cascading outages.

Experimental Results

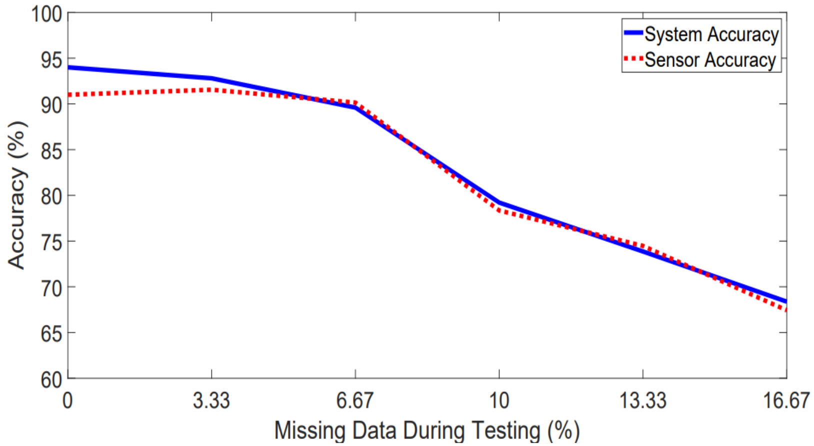


Fig. 1 Sensitivity of event identification accuracy to the size of missing data.

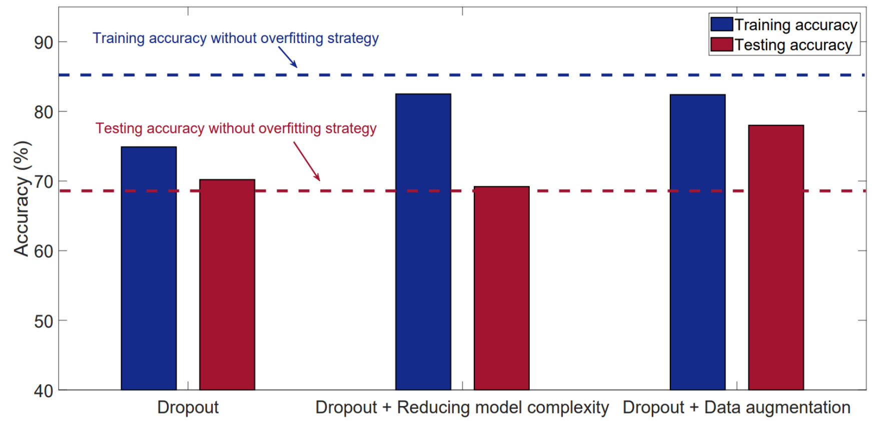


Fig. 3 Comparison of different overfitting strategies.

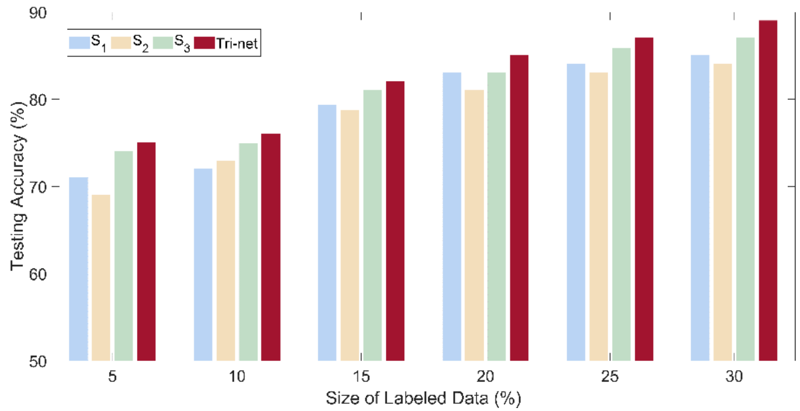


Fig. 2 Sensitivity of event identification accuracy to the number of labeled events.

- Our proposed methods aim to address three common challenges encountered when using PMU data and machine learning techniques for event identification: (I) robustness (Fig. 1) (II) data scarcity (Fig. 2) (III) reliability (Fig. 3).

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Experimental Results

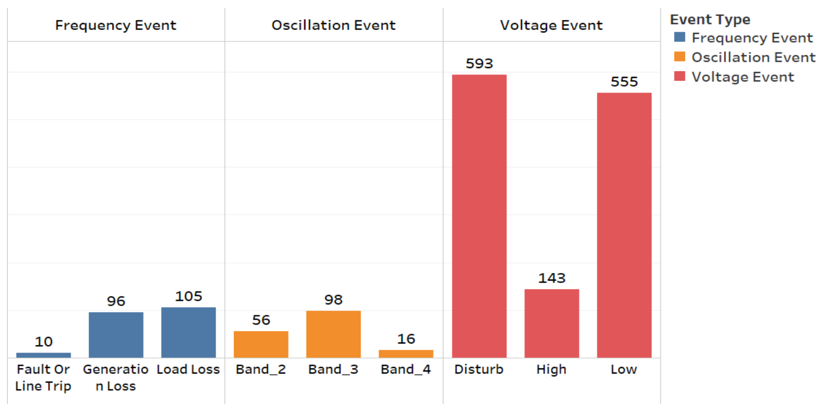


Fig. 4 Experimental summary for system A using EPG software.

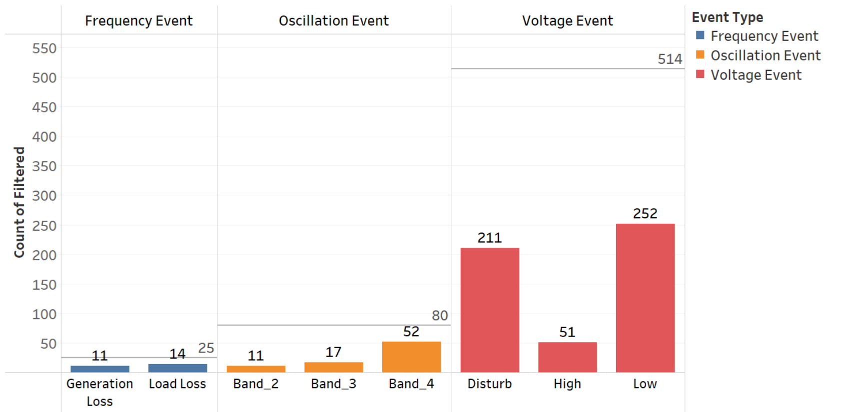


Fig. 6 Experimental summary for system C using EPG software.

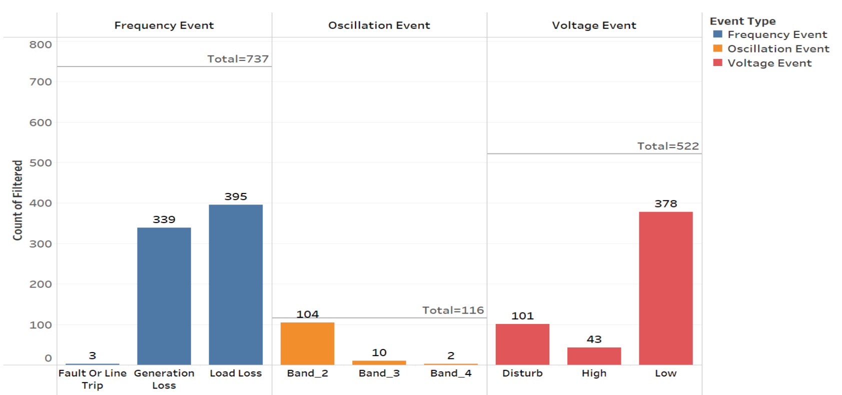


Fig. 5 Experimental summary for system B using EPG software.

- Our experiments aim to use commercial software to update recorded events and discover additional events in the training and testing dataset.
- Unlike machine learning models that rely on the event labels, commercial software is based on default criteria to identify events.

Technical Accomplishments

- Comprehensive data quality assessment for the entire dataset (including training dataset and testing dataset) using PMU status flags information.

- Developed a robust learning-based event identification model to introduce robustness to the online PMU data dimensional imbalance problem.

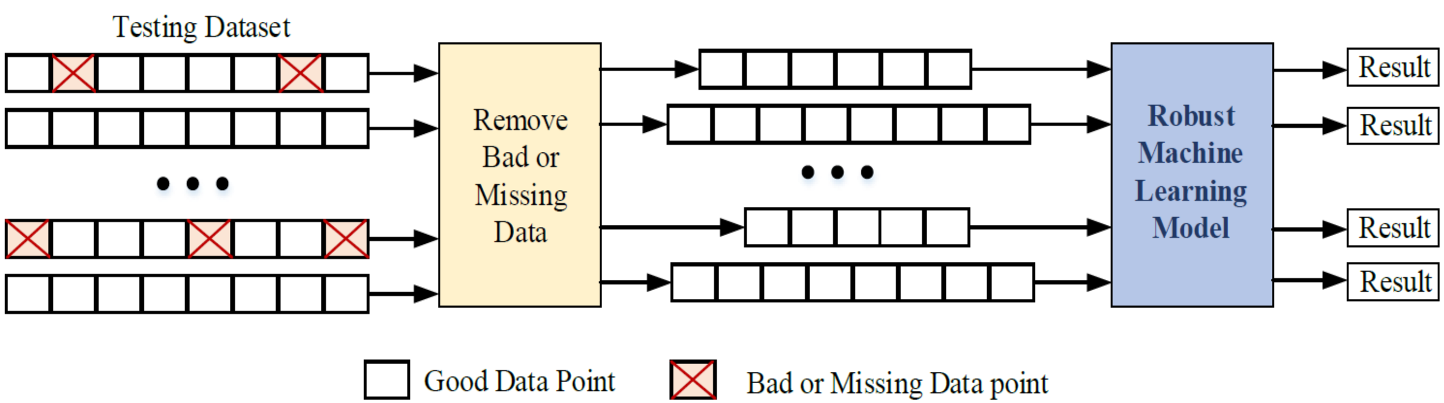
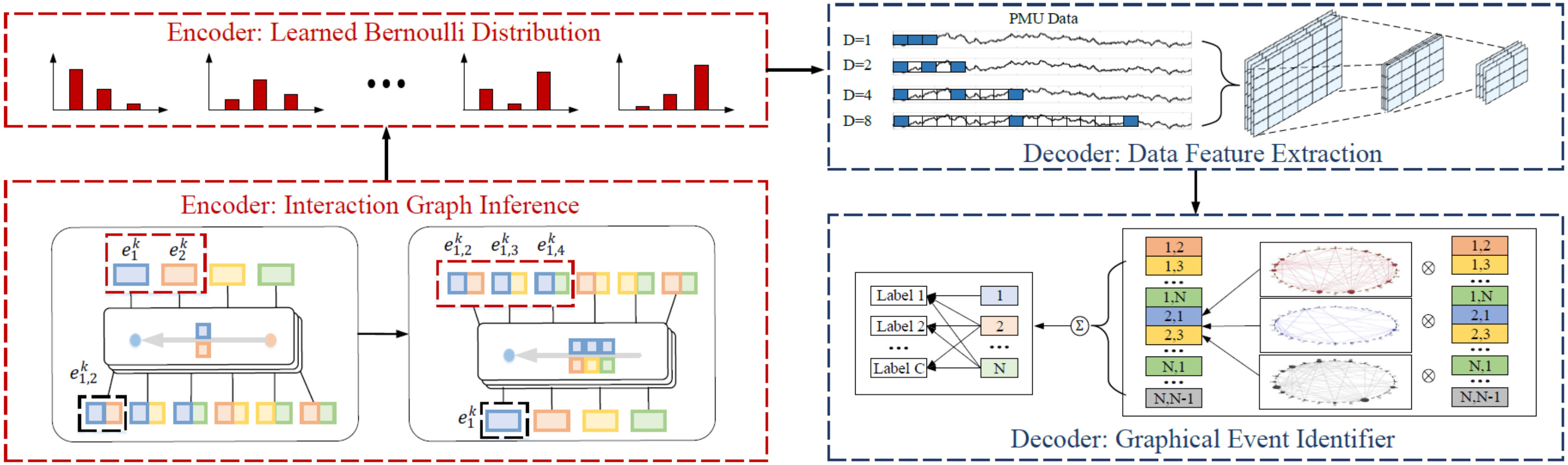


Fig. 7 Illustration of our method for solving dimension imbalanced problem.

Technical Accomplishments

- Developed a semi-supervised event identification model to improve the performance of the event identification model when trained with a limited number of recorded events.

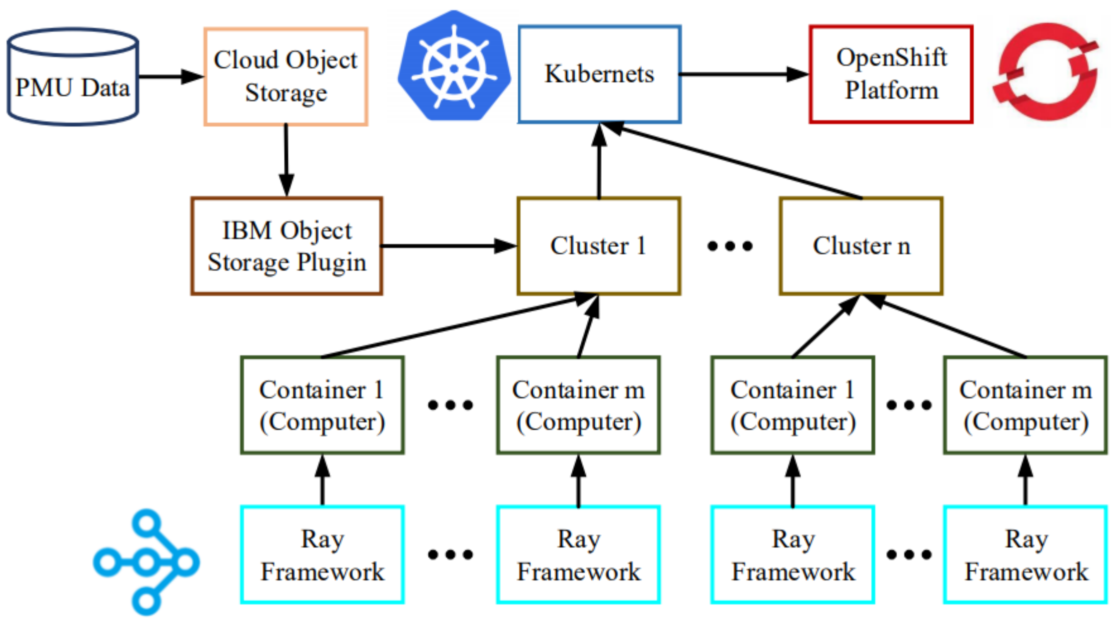
- Developed a method to learn latent interaction graph jointly with event identification task using deep graph learning.



Technical Accomplishments

- Identified cascade failure and developed a near-miss search method using support vector machine.

- Developed a cloud framework for parallel and distributed PMU data processing using Ray and OpenShift Kubernetes.



Value of Work

What additional benefit can this bring to the utilities?

- Our proposed deep learning-based event identification models can be treated as supervisory monitoring modules for utilities. When SCADA is dysfunctional, our data-driven models can continue to work, thus maintaining partial system awareness.

- The high-value use cases for our algorithm are that our methods can still give meaningful conclusions when utilities have imperfect measurement data and a small number of manually recorded event labels.

- Our works provided more precise time stamps for event logs, which can be utilized for future event identification method development and validation.

Value of Work

This dataset included data from multiple utilities spanning the interconnections. Based on this experience, what can utilities gain by sharing data with each other?

- Utilities can get more comprehensive PMU data analysis as well as more reliable event identification models by sharing data with each other. After parameter fine-tuning, a good model should be able to get good results in multiple interconnected data.

What are the limitations of working with anonymized data? Is it worthwhile for utilities to share anonymized data?

The limitations of working with anonymous data are mainly in the inability to access and validate some system-related information, such as system topology and the location of events. Sharing anonymized data is still worthwhile because real-world data is one of the foundations for building a reliable data-driven model

Readiness for Commercialization

What would be the next steps for making the results from the projects available to use by utilities?

- The utilities can verify about the suspicious events (i.e., events recorded on the event logs without event features in the PMU data) and additional events (i.e., events that do not appear in the event logs) that we discovered.

If you could put a readiness number on your product, 0 being fundamental early-stage research and 10 being commercially viable for sale as a working product, where you rate your product? How do you anticipate transitioning your research to tools that are available to utilities? In the near-term?

- We would give our works a 6 out of 10. Our works take into account many practical challenges. However, the quality of the event labels of the current dataset is not sufficient. Therefore, the performance of these methods will require more discussion and validation.

Being Ready for ML & BD Analytics

Does off-the-shelf machine learning models achieve good performance for PMU data analytics?

- When the training and testing data is sufficient and perfect, off-the-shelf machine learning models can achieve acceptable performance for PMU data analytics. However, there is still much room for improvement

What are key challenges of AI/ML methods when it comes to analyzing power system data? How do you think those challenges can be addressed?

- The first challenge is how to deal with the special characteristics of power system data, such as high sampling rate, hidden interactions among PMUs, and event data scarcity. The second challenge is how to improve the interpretability of AI/ML models for power system applications. Solving these challenges requires a comprehensive understanding of both AI/ML and power systems. This enables the design of appropriate strategies for different challenges.

Being Ready for ML & BD Analytics

What additional data or change in data collection process could help in the development of ML models in the future? Labels are important for machine learning. What recommendations would you give a utility to make their event logs or other sources of labels useful for training AI/ML models?

- Additional location information of events and PMUs can help the development of future ML models.

- A uniform event label collection process will help to train AI/ML models. Thresholds of the data collection process can help us to evaluate the reliability of the labels as well as further classify the labels.

Being Ready for ML & BD Analytics

How important is it to synergistically combine machine learning models with power systems domain knowledge?

- Synergizing machine learning models with power system domain knowledge is very important. By simply applying off-the-shelf machine learning models, we cannot get a tool that is adequate to deal with the multiple complex scenarios of power systems.

What low-cost steps should utilities take now to make them ready for big data analytics 2 or 3 years from now?

- Low-cost steps could include further improving the quality of their own data and reaching out to the cloud computing.

Lessons Learned and Next Steps

What recommendations do you have for using the FOA 1861 dataset moving forward?

- The quality of the FOA 1861 dataset can be further improved by integrating everyone's results and performing cross-validation. Considering the size of the database, a low-cost cloud computing platform will improve the applicability of the FOA 1861 dataset

Was it difficult to distinguish significant threats from the common disturbances that happen every day? How did you address this difficulty?

- Yes, the reason for this difficulty is that the labels are now ambiguous. We studied the start-up time gap between consecutive line outages to identify cascading failures from the event logs.

Lessons Learned and Next Steps

How to improve the interpretability of data-driven models for power system event detection and classification?

- Complex models are less interpretable because their relationships are generally not summarized in a concise manner. Balancing the trade-off between model complexity and interpretability is one of the keys to improving the interpretability of data-driven models for power system event detection and classification.

- Human interpretation can be enhanced if visual and interactive diagrams and figures are used for the purpose of explaining the data-driven models for power system event detection and classification, such as a clear visualization of event features in convolutional layers.

Lessons Learned and Next Steps

What would be next steps for research of related technologies? As a result of this work, how do you foresee ML/AI being used on the grid today and in the near future?

- The next step in related technologies may be how to elegantly handle multiple challenges at once. Meanwhile, it is a big challenge to ensure the generalization of data-driven event identification and classification models.

- The widespread high-resolution sensors has enabled utilities to collect an unprecedented amount of demand-side and system-side data that facilitates the transition to a data-enabled modernized power system. It is foreseeable that ML/AI will be used more and more on the grid. The key is how to design practical models that fit the goals of the smart grid.

Publications List

- Our paper entitled "***Learning-Based Real-Time Event Identification Using Rich Real PMU Data***," has been accepted by IEEE Transactions on Power Systems.
- Our paper entitled "***Learning Latent Interactions for Event classification via Graph Neural Networks and PMU Data***" has been accepted by IEEE Transactions on Power Systems.
- Our paper entitled "***Discrete Graph Structure Learning for Forecasting Multiple Time Series***" has been accepted by International Conference on Learning Representations (ICLR) 2021.
- Our paper entitled "***Graph-Augmented Normalizing Flows for Anomaly Detection of Multiple Time Series***", has been accepted by ICLR 2022.
- Our paper entitled "***Federated Inference through Disambiguating Local Representations and Learning a Consensus Graph***", is in preparation.
- Our paper entitled "***A Safe Semi-supervised Learning Framework for Event Identification Using PMU Data***" is in preparation.