FOA 1861 FINAL PROJECT BRIEFING
BIG DATA ANALYSIS OF SYNCHROPHASOR DATA

Big Data Synchrophasor Monitoring and Analytics for Resiliency Tracking (BDSMART)

Presenter: Mladen Kezunovic, PI

Texas A&M Engineering Experiment Station (TEES)

m-kezunovic@tamu.edu

Project Partners:

Zoran Obradovic (Co-PI), Temple University

Yi Hu (Co-PI), Quanta Technology

NASPI Meeting

April 12-14, 2022

Project Overview

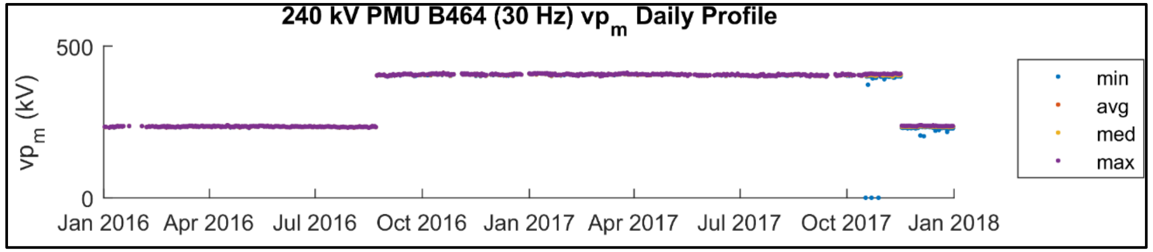
- Project Objectives:
  - Understand data properties
  - Set up data for model development and validation efforts
  - Develop models to detect, characterize and classify events
- Technical Approach:
  - Deploy the expert domain skills and ML/AI approaches to discover and utilize knowledge for development of data models
- Significance and Impact:
  - Develop useful models, offer recommendations for best practices, and facilitate initiation of new standardization efforts

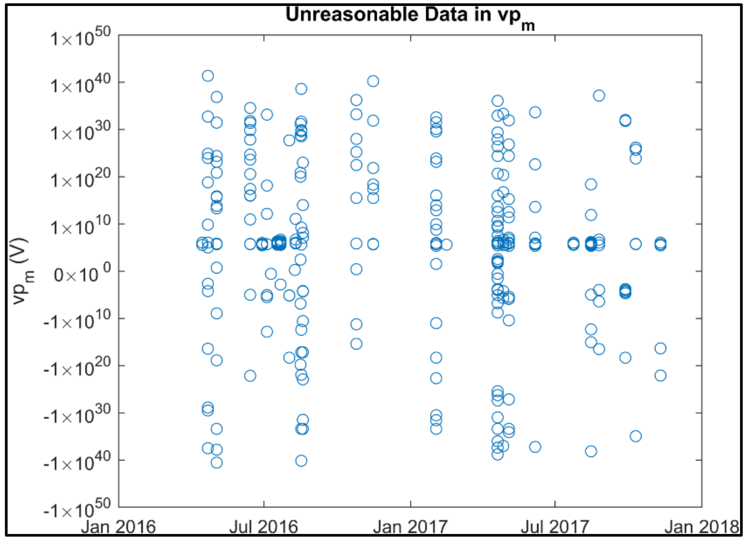
Experimental Results Background

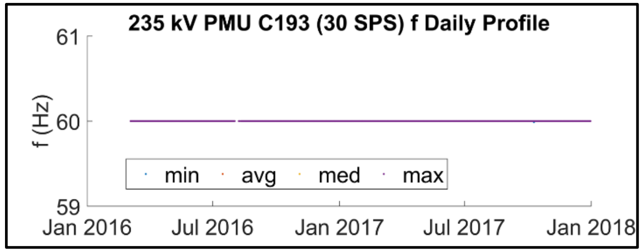
- Tasks 1&2: Project management and planning, and Data platform selection and configuration:
  - Achieved tight Team coordination through bi-weekly meetings
  - Implemented HPRC setup at A&M with access from Temple, and waveform viewing and simulation setup at Quanta
- Tasks 3&4: Data quality assessment and initial ML/AI event classification, and Event analysis:
  - Performed visual data inspection and data/labels correction initially
  - Facilitated automation of event detection using a v/f rectangle feature
- Tasks 5&6: Event labeling/categorization, and event prediction:
  - Used improved labels and applied sliding data-window to characterize event sets and types
  - Used variety of ML/AI techniques to account for information loss due to imprecise labels
- Tasks 7&8: Evaluation, and Preparation of required deliverables
  - Defined evaluation criteria and metrics for model development, and testing, and applied to ICB
  - Prepared a report for ICB and now running experiments to include reports on ICC and ICA

Experimental Results – Data Quality

- Inconsistent voltage level
- Missing data
- Unreasonable data:
  - Extreme values
  - Flat 60 Hz
- Erroneous time tag
- Artificially altered data
- Data duplicates







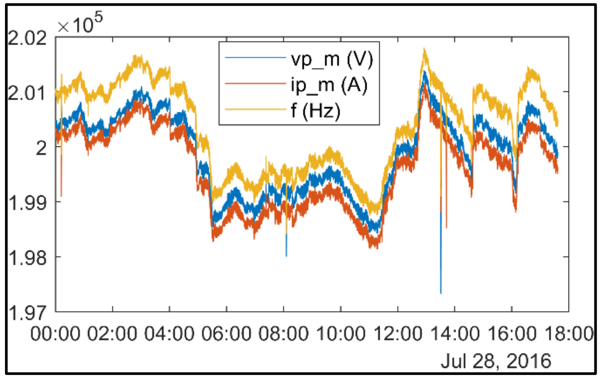
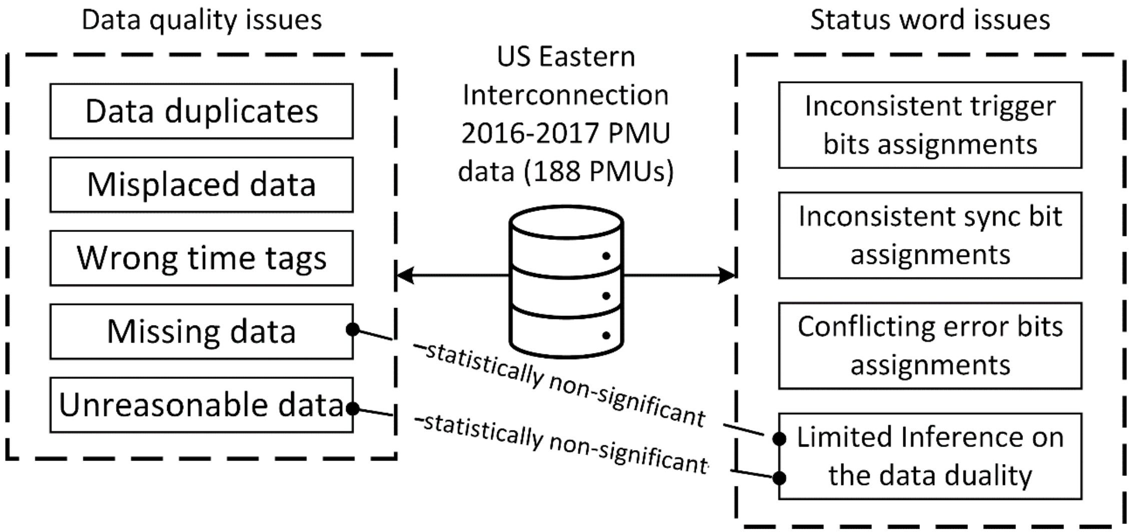


TABLE II. WRONG TIME TAG IN PMU C569'S DATA.

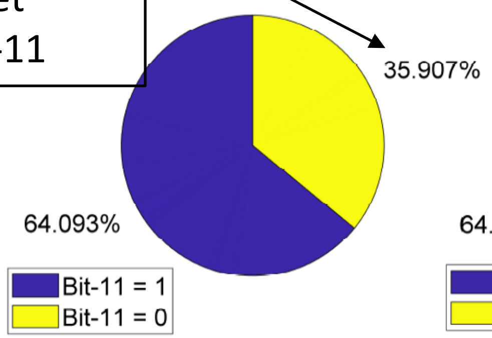
UTC	vp_m	vp_a	ip_m	ip_a	f	df
13:15:17.76667	208258.95	49.00	387.79	53.00	60.02	0.002
13:15:17.76676	NaN	NaN	NaN	NaN	NaN	NaN

Experimental Results – PMU Status Word

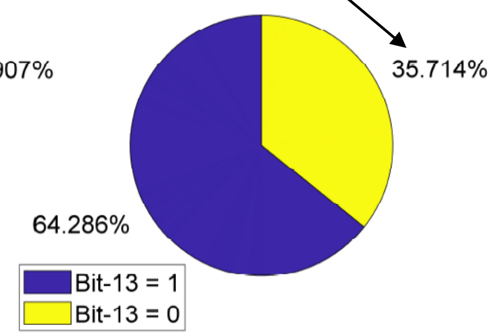
- PMU status word cannot be used as a reliable data quality indicator
  - Does not always match actual data quality.
  - Inconsistencies in PMU status bits assignments

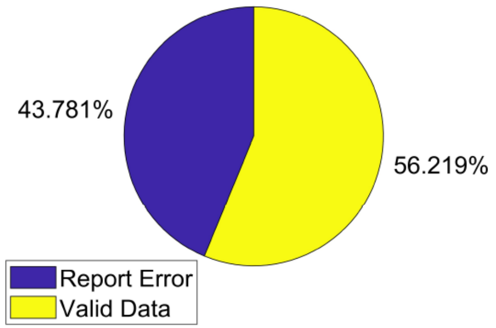


When trigger condition bits are set, 36% of the PMU don't set triggered bit-11



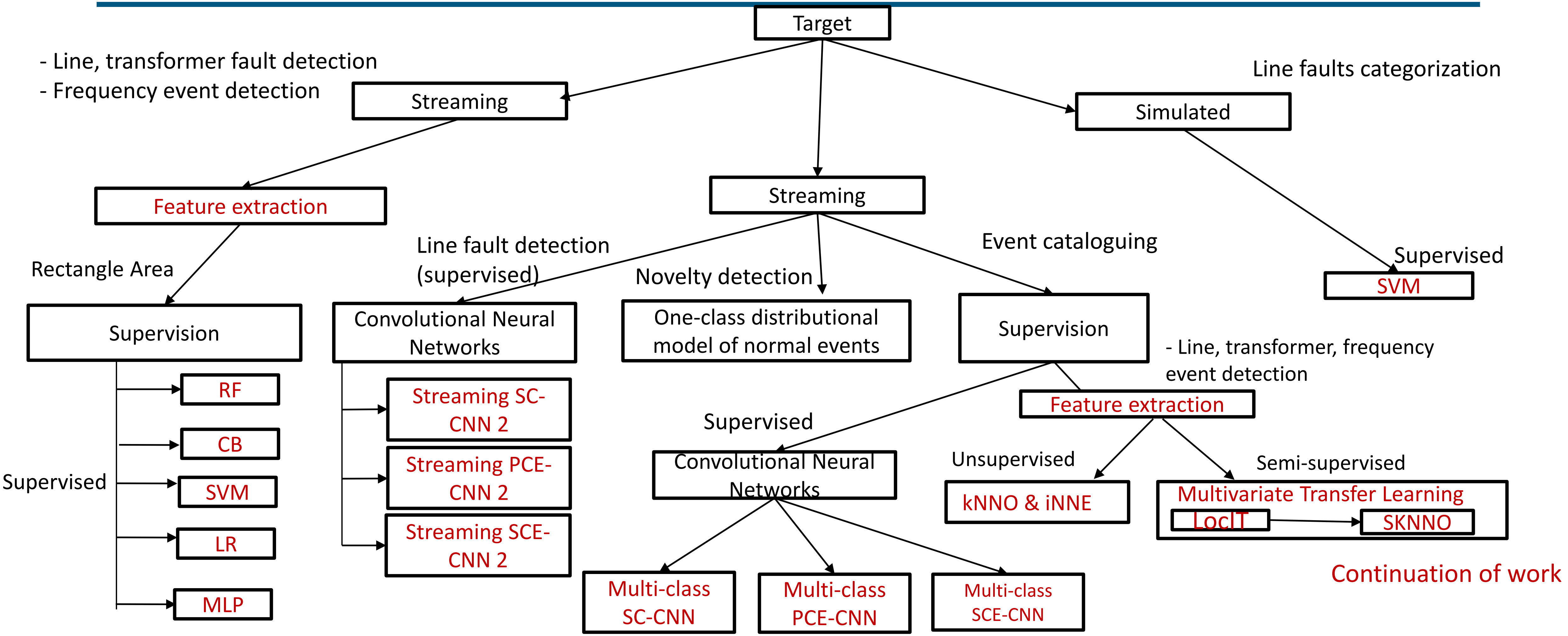
When clock is unlocked, 36% of the PMU doesn't flag bit-13 as out-of-sync





When PMU is sync, 56% of the PMU doesn't set error bits indicating valid data

Experimental Results

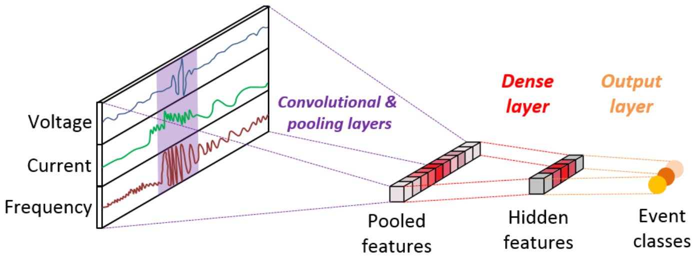


Technical Accomplishments

Problem formulation: Given a **signal segment** $\mathbf{s}(t-\Delta, t+\Delta) = [\mathbf{s}^{(1)}(t-\Delta, t+\Delta), \ldots, \mathbf{s}^{(M)}(t-\Delta, t+\Delta)]$, from *multiple anonymized PMUs* **predict event type** $y \in \{0, \ldots, C\}$ that occurred at $[t-\Delta, t+\Delta]$ by learning from *scarce observations* and *low precision labels*.

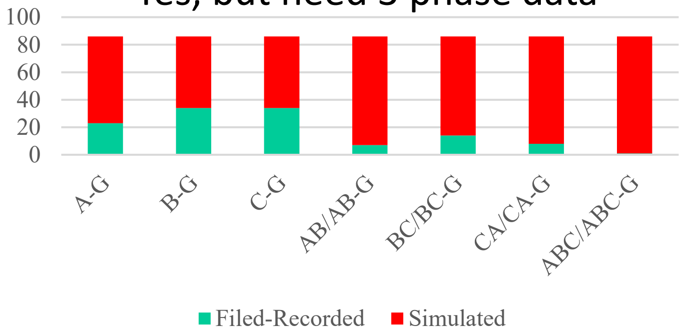
Q1: Can feature learning be automated?

Yes – Multi-channel filtering by CNN



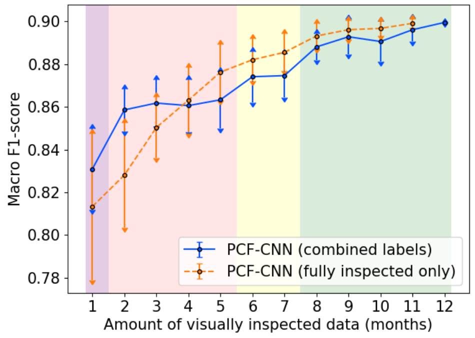
Q3: Can models be improved by using PMU data from simulations?

Yes, but need 3 phase data



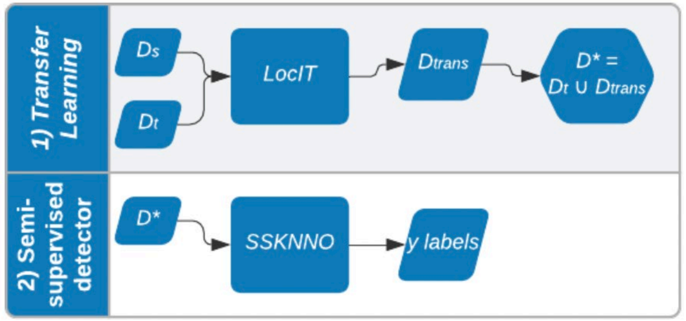
Q2: Should models learn from more data or from better data?

Use both if data is small



Q4: Can relevant labeled PMU data from a related task be used for learning on a new tasks?

Yes - transfer learning



Technical Accomplishments

- Developed a supervised learning method for local and system-wide event detection in power grids using sparsely placed PMUs
  - Problem: Develop a scalable automated event detection systems that doesn't rely on extensive manual study of data and feature engineering. Utilize a sparse set of PMUs that don't necessary cover all geo areas. Handle noisy data and unknown event locations.
  - Accomplishments: Automated data preprocessing steps and three CNN based detection models were introduced. Data from western interconnection were used (2016 for training and 2017 for testing). Robust event detection is achieved in multiple settings, but multi-channel hierarchical CNNs outperformed alternatives. Curating event logs leads to increased detection accuracy. Fully inspecting at least two months of data is suggested.
- Developed a line fault, frequency, and transformer event detection method based on transfer learning techniques
  - Problem: Detect events based on minimal labeled time windows by leveraging related labeled instances from another domain without relying on event logs of PMU data.
  - Accomplishments: The transfer learning method yielded ~13% improvement in AUROC when compared to supervised learning algorithms based on only 20 labeled time windows. 2-seconds time window yielded an approximately 7% increase in AUROC compared with 1-minute windows.
- Developed line faults classification using machine learning on three phase voltages
  - Problem: Classify line faults when field-recordings have insufficient number of observations of certain type (e.g. PP, PPG,3P, and 3P-G faults).
  - Accomplishments: A classification model trained on integrated simulations and field-recorded data resulted in 98.5% accuracy. This is a significant improvement over 87.17% accuracy obtained by relying on the field-recorded data alone.
- Develop voltage level-aware CNN classifier based on sliding window technique
  - Problem: Classify events as normal, line, and frequency events using PMU measurements split into voltage levels (134kV, 240kV, 300kV, 500kV).
  - Accomplishment: A 30s sliding window is subsampled from the preprocessed data, and those subwindows are used as CNN classifier inputs. A split into voltage levels is beneficial when applying signal summarization through a Soft-Dynamic Time Warping.

Value of Work

- Establishing facts:
  - Even now, let alone 5 years from now, collecting PMU data may have diminishing returns unless the analysis is automated
  - While automating the analysis of historical data has value, real benefit comes from predicting occurrence and mitigating impacts of undesirable events in real-time
- Offering Recommendations:
  - While utilities can gain by sharing data with each other, the real value is in sharing and following best practices in recording its own data, and then preserving it
- Facilitating standardization work:
  - Developing and adopting standardized approaches on PMU setting flags, the exact meaning of the error bits, and a common format for event labeling

Readiness for Commercialization

- We have developed data model prototypes that can be tailored to specific applications: event detection, fault , and system-wide analysis
- Our prototypes are at the TRL 6: System/subsystem model or prototype demonstration in a relevant environment.
- To make our protypes commercially viable for sale as a working product, we are in the early stages of product development
- The transitioning of our research to tools that are available to utilities would require access to more accurately labeled data with physical model information, and definition of more specific goals of the tools

Being Ready for ML & BD Analytics

- Off-the-shelf machine learning models, while certainly a good starting point for education and training, are not going to achieve good performance for PMU data analytics without tuning
- The key challenges of ML/AI methods when it comes to analyzing power system data is in automating the data labeling, including time-stamping, as well as in capturing long data history
- When focusing on PMU data and event logs:
  – Improving data quality and recording practices could help in the development of ML/AI models in the future.
  – Data labeling should be done not only based on SCADA data but also based on data from other recording systems, including GPS time-stamps
  – It is essential to synergistically combine machine learning models with power systems domain knowledge since data-based models, as powerful as they are, will not be sufficient.
- The low-cost steps for utilities to take now to make the ML/AI approaches ready for big data analytics 2 or 3 years from now is to amend and open data sets for ML/AI experts to use.

Lessons Learned and Next Steps

- Improving data labeling, data quality and recording practices (three-phase vs positive sequence) is essential for future ML/AI applications in the utilities
- Providing power system topology and PMU placement helps in distinguishing power system events and assessing their importance/impact
- Using synthetic data has limited value except for the fault studies where the events are local, highly distinguishable, and resemble actual events closely.
- Data management of large data sets of streaming data is an expensive effort and requires new data management, data wrangling, and data viewing tools
- Enhancing PMU data with data from other utility recording devices/systems (DFRs, DPRs, SCADA), and with weather data can produce significant benefits
- Automating event detection and analysis, evolving from a posterior (historical) to a priori (predictive) formulation, requires further ML/AI research

Publications

Journals

- M. Alqudah, M. Pavlovski, T. Dokic, M. Kezunovic, Y. Hu, Z. Obradovic, “Convolution-based Fault Detection Utilizing Timeseries Synchrophasor Data from Phasor Measurement Units,” IEEE Transactions on Power Systems, 2022. DOI: 10.1109/TPWRS.2021.3135336.
- M. Pavlovski, M. Alqudah, T. Dokic, A. Abdel Hai, M. Kezunovic, Z. Obradovic, “Hierarchical Convolutional Neural Networks for Event Classification on PMU Data,” IEEE Trans. on Instrumentation and Measurement, Vol. 70, pp. 1-13, no. 2514813, 2021.
- A. Abdel Hai, T. Dokic, M. Pavlovski, T. Mohamed, D. Saranovic, M. Alqudah, M. Kezunovic, Z. Obradovic, “Transfer Learning for Event Detection from PMU Measurements with Scarce Labels” IEEE Access, Vol. 9, 127420 – 127432, September 2021.

Conferences

- M. Kezunovic, Z. Obradovic, Y. Hu “Automated System-wide Event Detection and Classification Using Machine Learning on Synchrophasor Data,” CIGRE General Session, Aug. 28-Sept 2, Paris, 2022.
- M. Kezunovic, Z. Obradovic, Y. Hu “Use of Machine Learning on PMU Data for Transmission System Fault Analysis,” CIGRE General Session, Aug. 28-Sept 2, Paris, 2022.
- Z. Cheng, Y. Hu, Z. Obradovic, M. Kezunovic, “Using Synchrophasor Status Word as Data Quality Indicator: What to Expect in the Field?”, IEEE Smart Grid Synchronized Measurement and Analytics Conference, SGSMA 2022, Split, Croatia, May 2022
- T. Dokic, R. Baembitov, A. Abdel Hai, Z. Cheng, Y. Hu, M. Kezunovic, Z. Obradovic, “A Single-Feature Machine Learning Method for Detecting Multiple Types of Events from PMU Data,” IEEE Smart Grid Synchronized Measurement and Analytics Conference, SGSMA 2022, Split, Croatia, May 2022
- H. Otudi, T. Dokic, T. Mohamed, M. Kezunovic, Y. Hu, Z. Obradovic, “Line Faults Classification Using Machine Learning on Three Phase Voltages Extracted from Large Dataset of PMU Measurements,” HICSS-55 Conference, Hawaii, USA, January 2022.
- R. Baembitov, T. Dokic, M. Kezunovic and Z. Obradovic, “Fast Extraction and Characterization of Fundamental Frequency Events from a Large PMU Dataset Using Big Data Analytics,” HICSS-54 Conference, Hawaii, USA, January 2021.