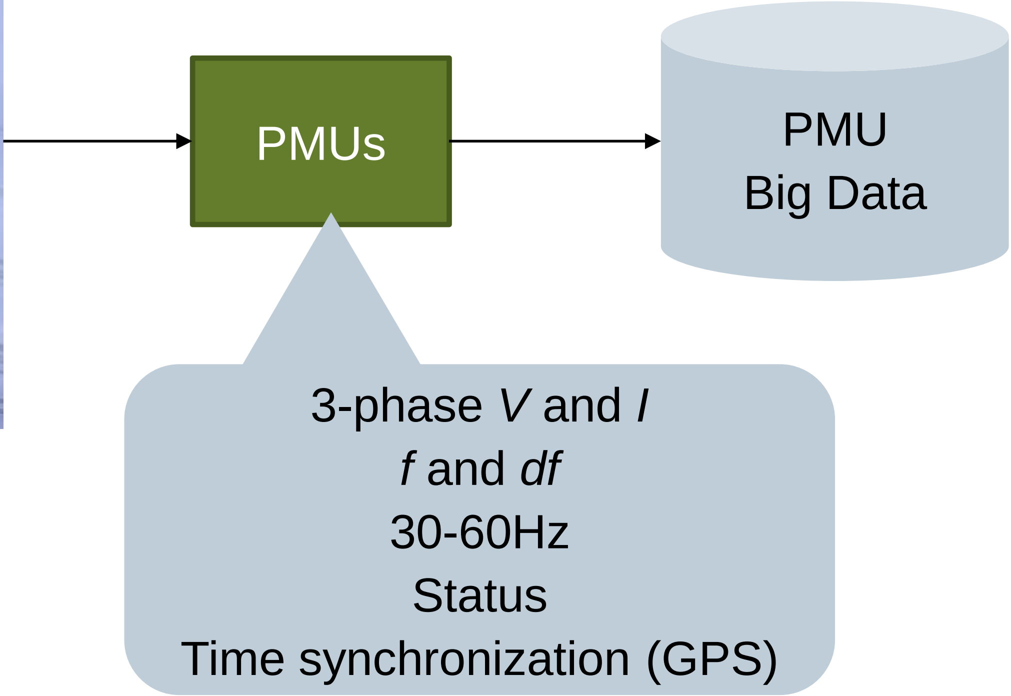
Big Data Processing for Power Grid Event Detection

Bruno Leao, Dmitriy Fradkin, Yubo Wang, Sindhu Suresh

IEEE Big Data Conference 2020, Second Workshop on Big Data Predictive Maintenance Using Artificial Intelligence (BDPM-AI 2020)

Introduction



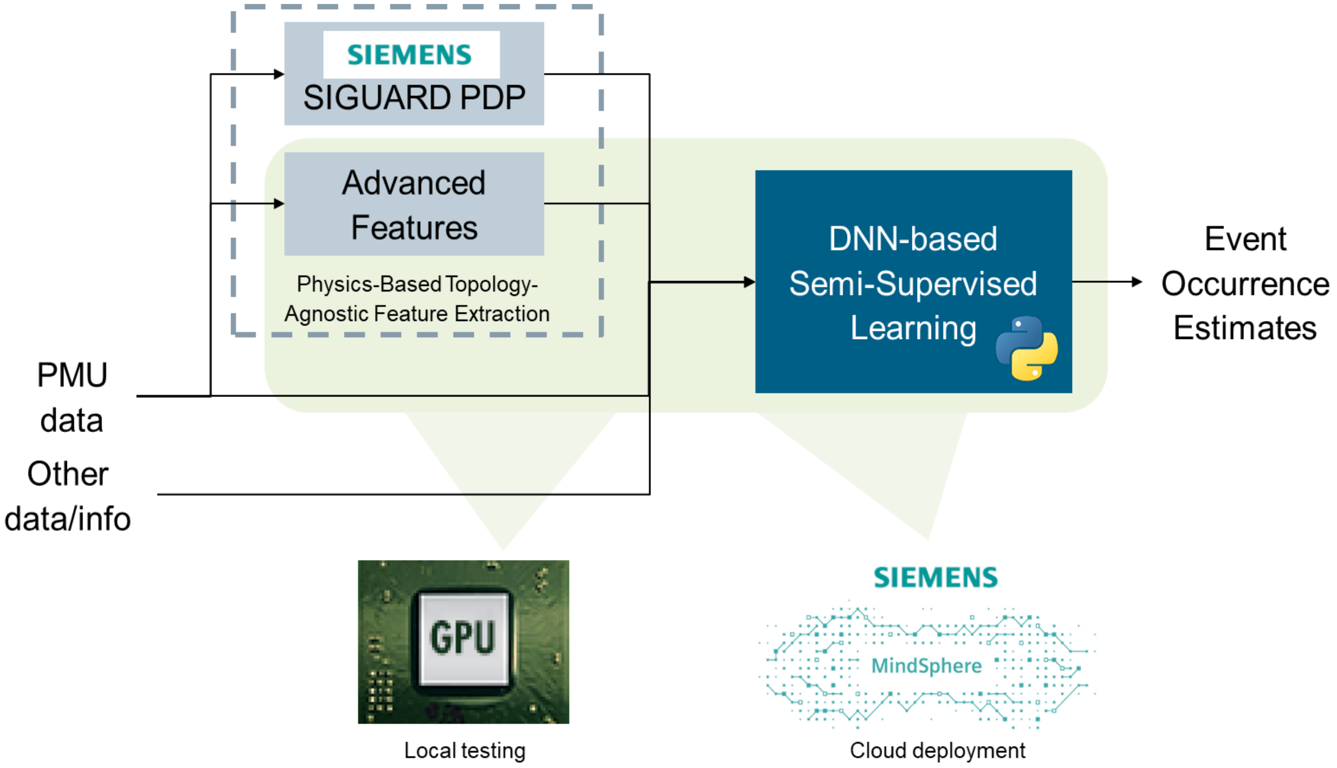
PMUs

PMU
Big Data

3-phase *V* and *I*
f and *df*
30-60Hz
Status
Time synchronization (GPS)

Introduction

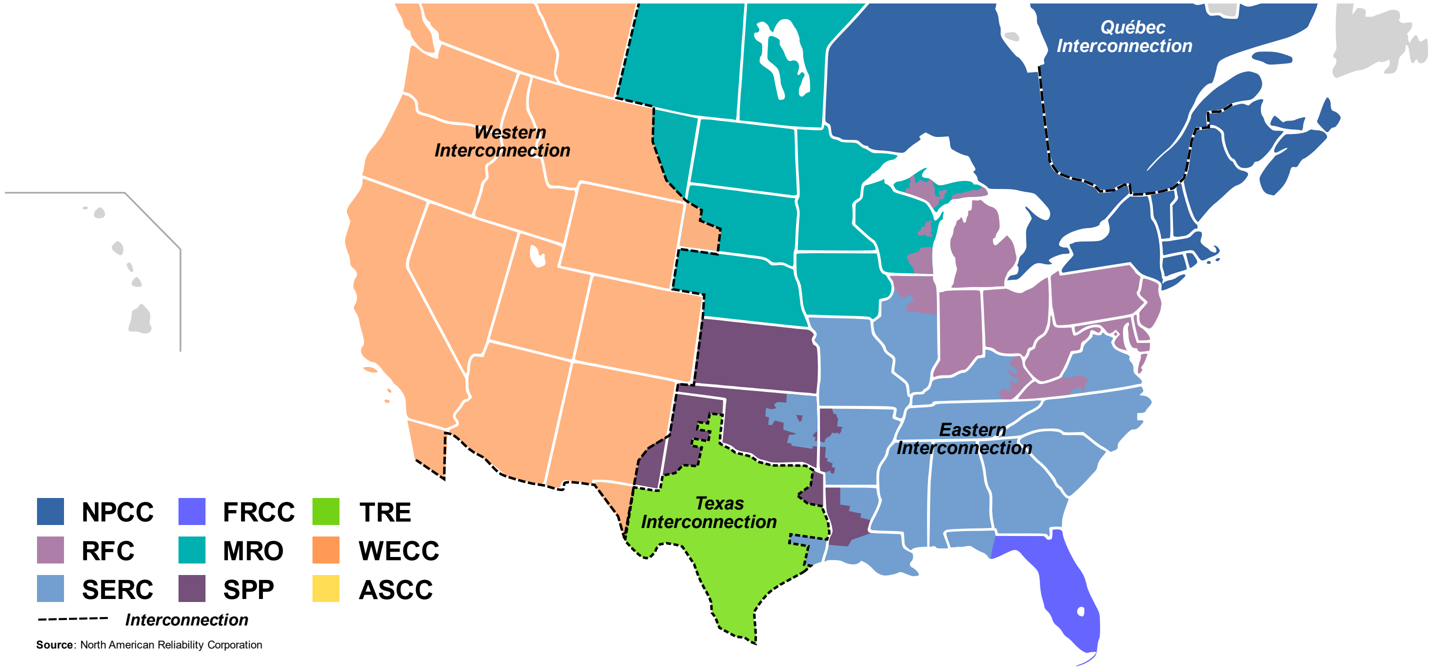
MindSynchro: DOE FOA 1861


SIEMENS
SIGUARD PDP
Advanced Features
Physics-Based Topology-Agnostic Feature Extraction
DNN-based Semi-Supervised Learning
Event Occurrence Estimates
PMU data
Other data/info
GPU
Local testing
SIEMENS
MindSphere
Cloud deployment

Big Data and Processing

SIEMENS

Ingenuity for life



Source: North American Reliability Corporation

Interconnection	A	B	C
Actual IC	Texas IC	Western IC	Eastern IC
Start Date	2018-07-21	2016-01-01	2016-01-01
End Date	2019-08-24	2017-12-31	2017-12-31
PMU Number	215	43	188
Data Volume	~3TB	~5TB	~11.5TB
File Count	2576	4365	10496

Challenges about the Data

PMU Data quality:

- Missing values or missing complete measurements
- Duplicate rows
- Unaligned timestamps for different PMUs
- Overlap between files

“Label” information: Event logs

- Purpose of the log is not appropriate for labeling data
    - Label doesn’t reflect the underlying phenomena but the factor causing the event.
        - E.g. Categories “wind” and “animal” may both refer to short circuit or line down events with same/similar pattern
    - Duration may reflect the consequence of the event (e.g. blackout)
- No mapping to specific PMUs
- Manually defined and manually post-processed
- Large subset of events which are not of interest (e.g. planned events)

Computational Infra-structure

GPU Server:

- Processor: Intel® Xeon® Silver 4210, with 40 cores
- 196GB of RAM
- 4 NVIDIA Quadro RTX 6000 GPUs, each with 24GB of RAM
- 2TB NVMe drive for operating system
- 42TB of HDD space for data storage
- Ubuntu 18.04.2 LTS operating system

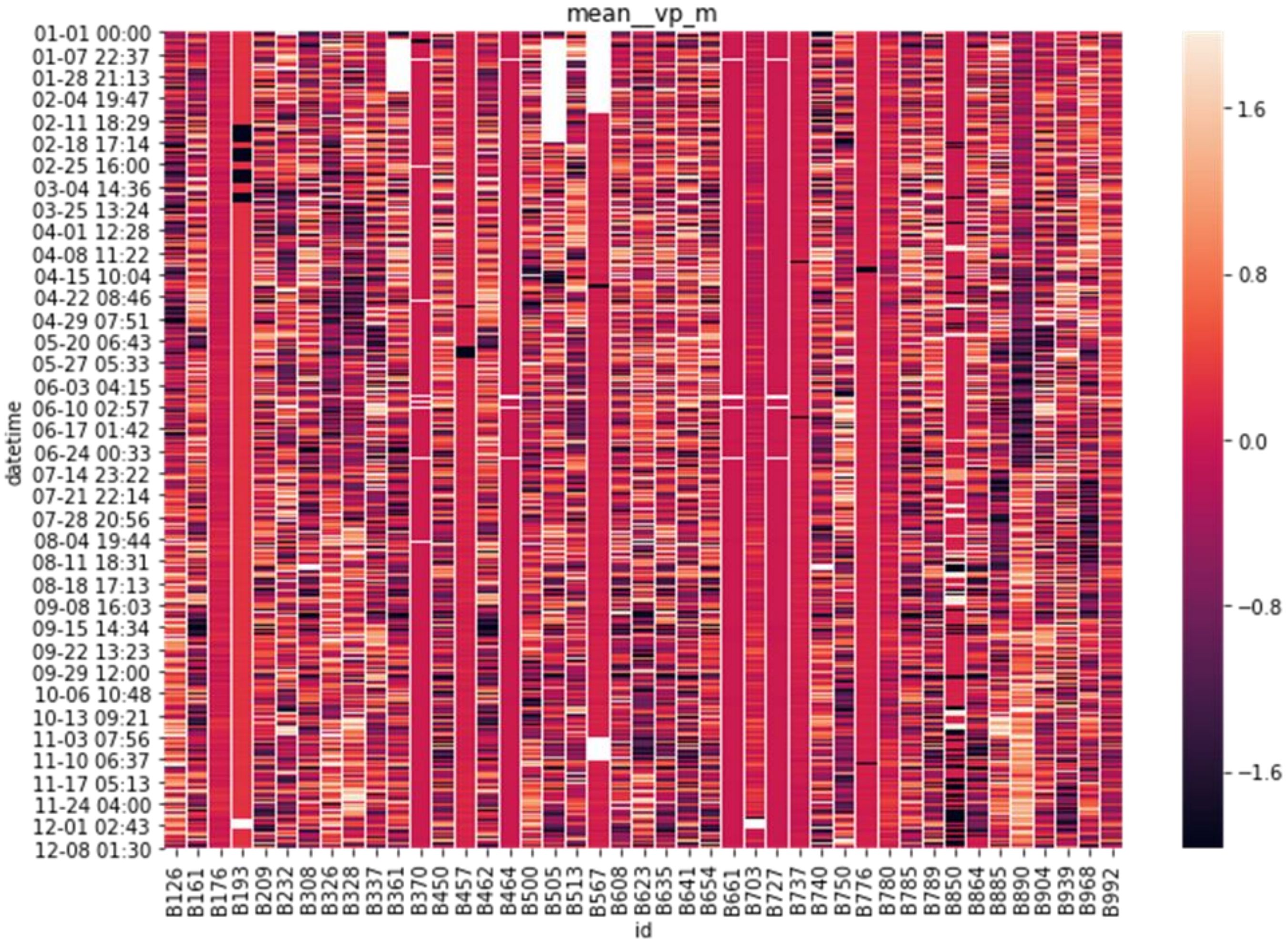
Pre-processing

1. loading original parquet file
2. converting numerical fields type (originally stored as strings)
3. converting empty values to null
4. removing columns which are all null
5. converting timestamp type (originally stored as strings)
6. indexing the data based on the timestamp
7. **calculating and saving statistics**
8. **synchronizing the data from all PMUs for each timestamp**
9. sorting by timestamp
10. saving the pre-processed data as a new parquet file

Status and outliers based data masks calculated as a separate pre-processing task

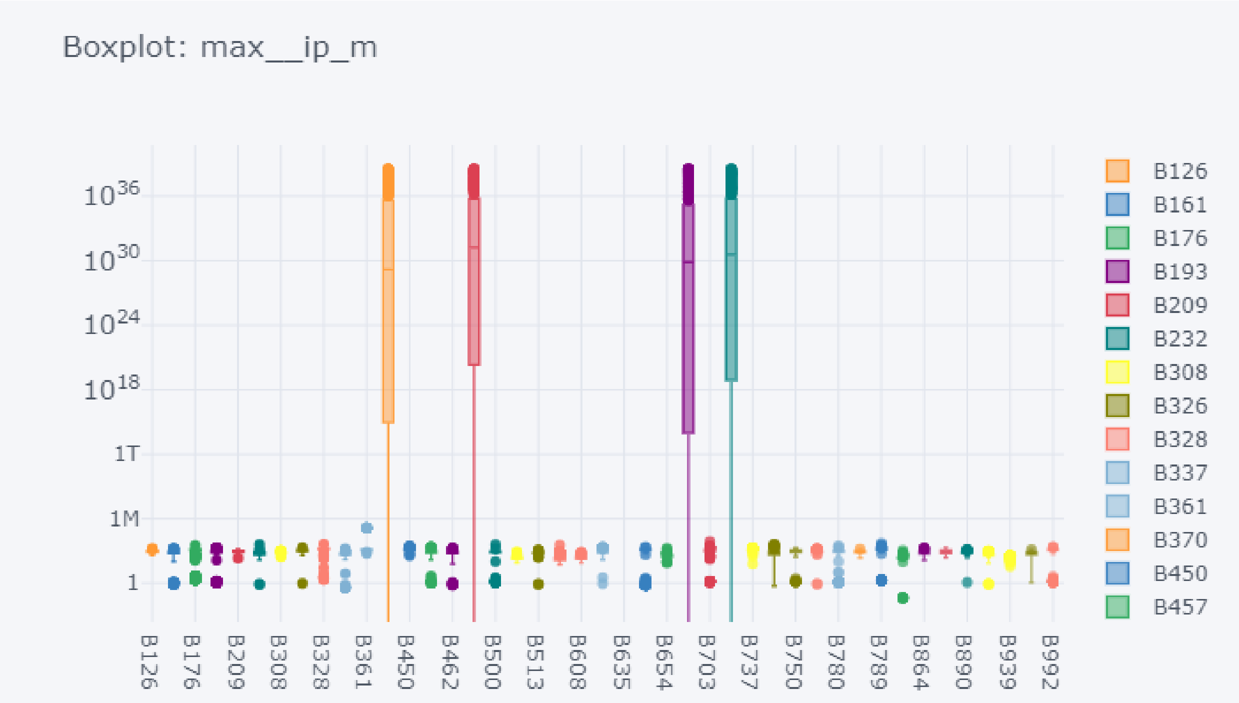
Leveraging GPU for pre-processing (RAPIDS cuDF)

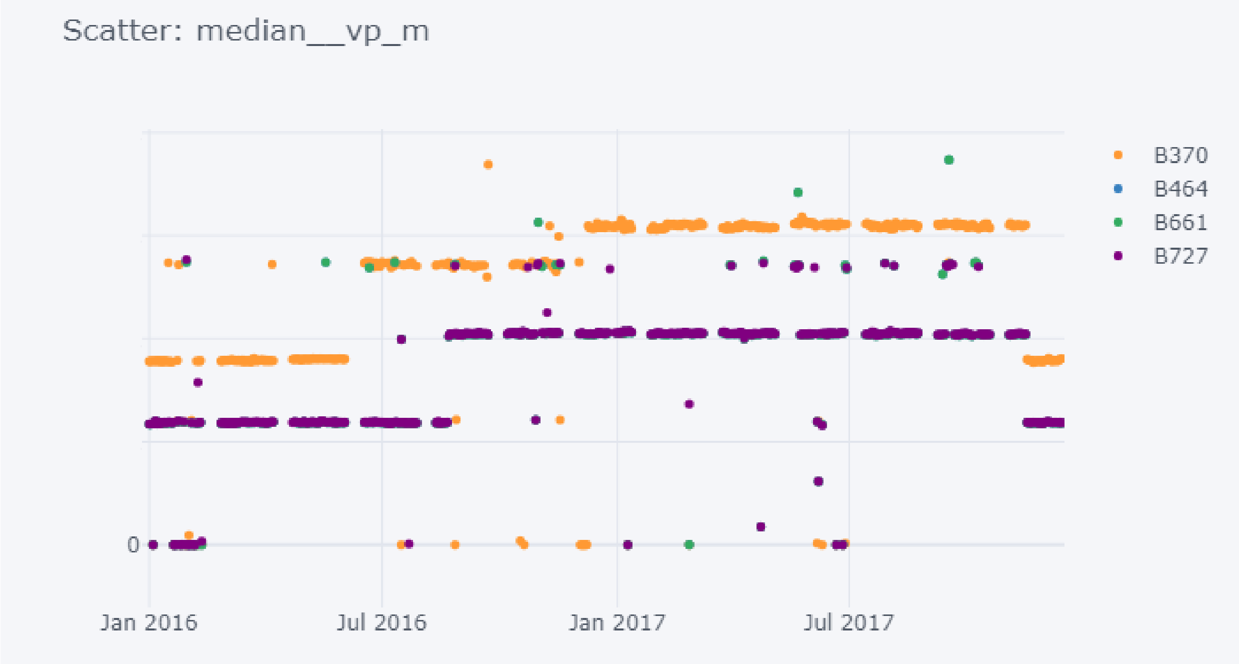
Data Visualization and Annotation



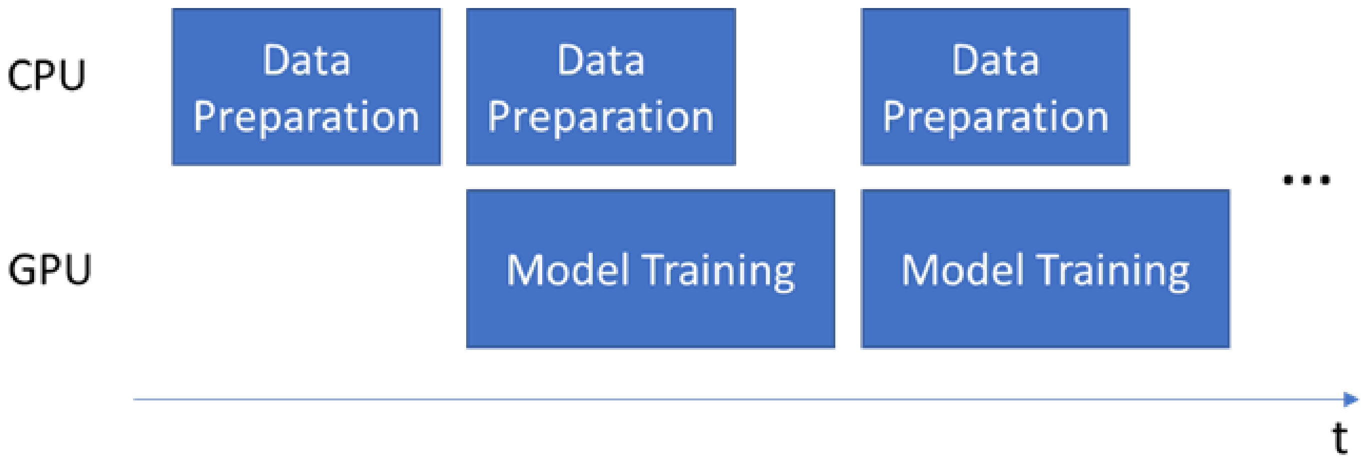
Unrestricted © Siemens AG 2020

Data Visualization and Annotation



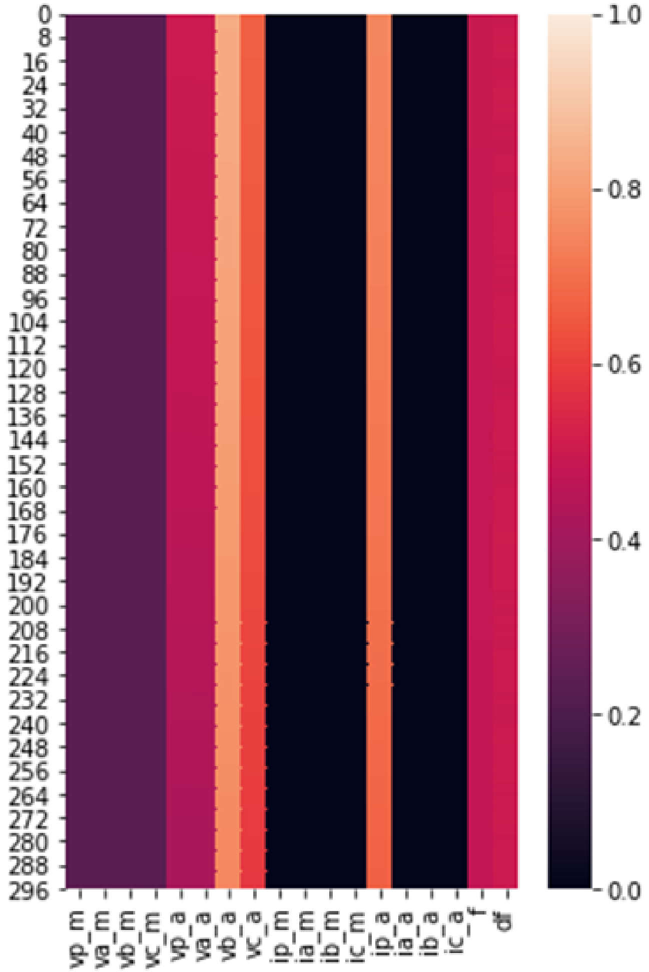


Machine Learning Development

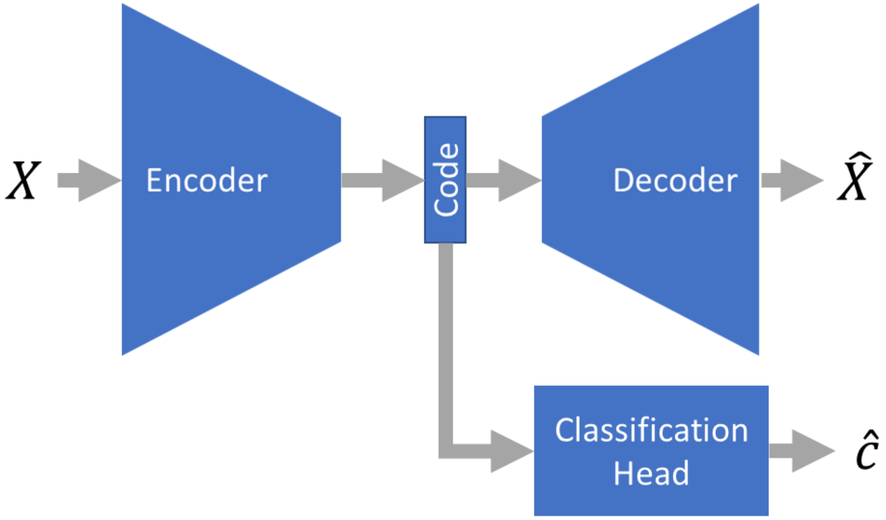


Batch preparation

1. load parquet metadata from pre-processed file and corresponding valid data mask
2. filter columns to specific PMUs if needed
3. load corresponding columns for data and mask
4. apply valid data mask to the data
5. eliminate regions of overlap with other files
6. **unwrap phase angles (optional)**
7. fill missing columns with token value
8. **resample to 30Hz**
9. **forward fill and then back fill missing values**
10. eliminate extra rows that won't fit in batch
11. **normalize values based on pre-defined ranges for each measurement**
12. reshape data to batch format
13. **adjust all angle values to the same reference in each batch (optional)**
14. replace token value from missing columns with zeros



Machine Learning Development



- Tensorflow 2
- Unsupervised learning dataset: over 370k samples (24 PMUs, 2 years of operation)
- Supervised learning dataset: short circuit, 221 labels from which 29 are positive
- Results: perfect accuracy

Conclusion and Next Steps

SIEMENS
Ingenuity for life

- PMU provides very rich big data
  - Well suited for detecting and categorizing relevant events
- Proper labels currently not available
- Challenges in dealing with real world big data
- Challenges in using the data for ML
- Sample ML results
- Next Steps:
  - Work in progress (project end expected by mid-2021)
  - Improvement in labels
  - Improvement in ML models
  - Insights from the complete dataset

Contact page

Bruno Leao

Siemens Technology

755 College Road East
Princeton, NJ 08540

USA

E-mail: bruno.leao@siemens.com

siemens.com