CHALLENGE 5B: IDENTIFICATION/CLASSIFICATION OF DANGEROUS EVENTS IN THE POWER GRID

INTRODUCTION

In this challenge you'll be dealing with one of the major problems that companies for high voltage power transmission are facing today.

Electricity runs from large powerplants through transmission powerlines and transformer stations all the way to factories, your home and other consumers. But every once in a while, a dangerous event happens on the side of either:

- Power Producers (unexpected events in large power plants)
- Transmission System (unexpected events in transformer stations and powerlines)
- Power Consumers (unexpected events at consumers).

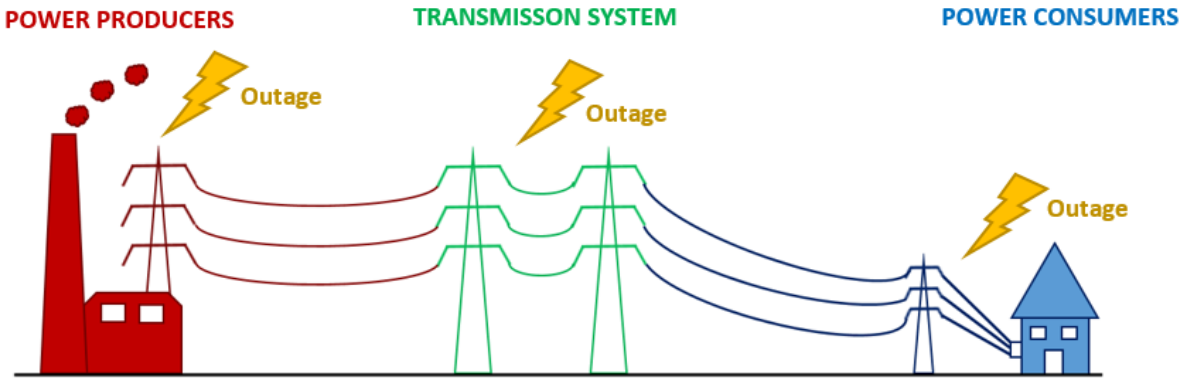

Figure 1: Flow of electrical energy from powerplants to consumers.

Examples of some of the dangerous events:

- Powerplant outage,
- Large consumer outage (e.g. big factory),
- Powerline outage,
- Transformer outage,
- Energy oscillations on powerlines.

DESCRIPTION

In some transformer stations in the Transmission System there are fast smart meters - sAMI which measure these events (measurement is taken every 20 ms constantly). These measurement devices are distributed across the Transmission System. Note, not all powerlines and substations are equipped with sAMI, so dangerous event can also happen at the unmeasured location. sAMI are time synchronized (by GPS clock), so every 20 ms a sample is taken at the same time in every sAMI in the grid.

Usually the dangerous events can be identified/classified by looking at the values of different variables that sAMI devices are measuring. The events are seen as transients of these variables. The measured variables are:

- voltage amplitude (of the busbar),
- voltage angle (of the busbar),
- current amplitudes (of each powerline in the transformer station),
- current angles (of each powerline in the transformer station),
- active power flows (of each powerline in the transformer station),
- reactive power flows (of each powerline in the transformer station),
- frequency (of the busbar),
- derivation of frequency (of the busbar).

Since the different parts of the power system are interconnected with powerlines **each sAMI in the system will measure the event**, that's why there are some transformer stations and some powerlines, which don't have sAMI. However, the further away that sAMI is from the event's location the less transient it will measure. There are many possible dangerous events that can happen in the system. Unfortunately, these events happen very rarely and they can be identified from the measurements only manually in post-mortem analysis and we want to automate this identification.

THE DATA AND THE OBJECTIVE

The objective of this challenge is to classify dangerous events and try to localize them on nodes from the available data so when the event in real world occurs the computer would know how to classify it and where to localize it.

In the folder *Data* you will find two subfolders:

- *RealMeasurements* (contains real-world measurements from sAMI devices taken when some dangerous event happened) – **11 events**.
- *Sims* (contains simulations of sAMI measurements at dangerous events) – **23 events**.

Each simulated/measured event is represented with a single *.csv* file, in which sAMI names are anonymized as NX (X being the number of the sAMI). In table 1 you'll see how the variables are represented in the *.csv* files.

Anonymized variables	Units	Meaning of anonymized variables
NX_u	[V]	Voltage amplitude at busbar (measured with sAMI X)
NX_au	[°]	Voltage phase angle at busbar (measured with sAMI X)

NX_in	[A]	Current amplitude at powerline n (measured with sAMI X)
NX_ain	[°]	Current phase angle at powerline n (measured with sAMI X)
NX_Pn	[W]	Active power at powerline n (measured with sAMI X)
NX_Qn	[VAr]	Reactive power at powerline n (measured with sAMI X)
NX_f	[Hz]	Frequency at busbar (measured with sAMI X)
NX_r	[Hz/s]	Derivation of frequency at busbar (measured with sAMI X)

Table 1: Meanings of anonymized variables.

Here is the structure of each *.csv* file:

- Columns: anonymized variables (first column is time).
- Rows: time series of simulated/measured variables.
- Resolution of time series: 20 ms.
- Delimiter: semicolon - ";"
- NaN values: only contained in *.csv* files in *RealMeasurements*.
- All sAMI devices are in every simulation, but not all sAMI devices are in every real measurement.

Which method you'll use for classification and machine learning is totally up to you. If you wish to do supervised leaning we're providing *table 2*, where there are identified events for every simulation *.csv* file.

Simulation *.csv* file	**Event classification**
Sim1	Power plant outage at N4
Sim2	Powerline outage between N1 and unmeasured N
Sim3	Powerline outage near to N5
Sim4	3 pole short circuit at unmeasured N
Sim5	3 pole short circuit at N1
Sim6	3 pole short circuit at unmeasured N
Sim7	3 pole short circuit at N2
Sim8	3 pole short circuit at unmeasured N
Sim9	3 pole short circuit at unmeasured N
Sim10	3 pole short circuit at unmeasured N
Sim11	3 pole short circuit at unmeasured N close to powerplant
Sim12	3 pole short circuit at unmeasured N close to powerplant
Sim13	3 pole short circuit at the the powerplant at unmeasured N
Sim14	Powerplant outage at unmeasured N
Sim15	Powerplant outage at unmeasured N
Sim16	Powerline outage near to N5
Sim17	Powerplant outage at unmeasured N
Sim18	Powerline outage between powerplants N10 and unmeasured N
Sim19	Powerline outage between unmeasured Ns (close to N7)
Sim20	Powerline outage between N1 and unmeasured N
Sim21	Powerline outage between unmeasured Ns
Sim22	Powerline outage between N2 and unmeasured N4
Sim23	Powerline outage between unmeasured Ns (close to N2)

In this case you can take the simulations for your training set and real measurements for validation set (**this is only our recommendation not a must**), to see if the classification of the real measurements produces any results (hint: *Sim1*, *Sim2* and *Sim3* are simulations of real events measured in *RealMeasurement1*, *RealMeasurement2* and *RealMeasurement3*, if you plot simulation against the real measurements, you will see some variations, this is due to electronic filters and noise in sAMI devices).