Exercise 1 - Initiate

David Fellner
01226059

October 2020

1 Project Idea

The project prospected is aiming at collecting a dataset related to electrical power grids. Furthermore, a baseline is to be established for working with the dataset collected. By collecting such a dataset the classification of abnormal behaviour of grid connected devices is to be made possible. Nowadays, electricity grid operators face many challenges connected to the fundamental changes the energy system is undergoing. Especially a high density of photovoltaic (PV) power generation has grave impact on a grid, as pointed out in [7]. Locally violations of the admissible voltage magnitude, the so called voltage band, are often the consequence, whereas the system frequency can be affected globally. To avoid such unfavourable effects, but without limiting renewable energy generation, control strategies are needed. Voltage regulation is regarded as the most important aspect in the integration of distributed generation in distribution networks [5]. This is implemented through grid supporting functionalities provided by the generation units. Amongst others, these range from curtailing the active power dispatched, to controlling the reactive power injection of generation units with inverters, which is commonly done via a local droop control. [6]

Grid simulation is to be used to synthesize data of scenarios in which devices that usually provide grid support functionalities experience malfunctions. These malfunctions cause them to stop providing this grid support. This ought to leave an impact on grid operational data. The grid participants should therefore act as life like as possible. This can be aided by applying real world load curves of household and energy dispatch patterns of generation units, such as PVs, to them. Data of regular operation and abnormal behaviour, as in the case of incorrectly parameterized control patterns, are needed to train a learning algorithm. Therefore, such data of low voltage grid participants, namely voltages, should be generated in this manner. The low voltage data should be synthesised at the devices in a way to mimic smart meter measurements. This can be achieved by adding noise to the generated, clean data. Figure 1 shows a schematic depiction of a low voltage grid setup that can be used for generating data: every terminal has a household load connected to it, but only a certain

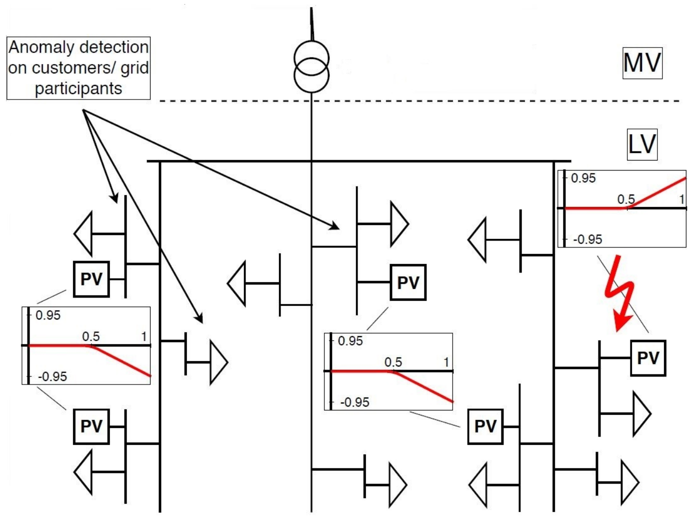

Figure 1: Schematic depiction of grid setup for data generation

share of the households are equipped with PV. All loads and PVs are connected to said terminals, which is in turn are connected to a feeder via a line. In the simulations, the data is recorded at these terminals, which would represent smart meters in the real world. All PV units are parameterized to follow the same control curve regarding reactive power dispatch. The curve applied is a cosϕ(P) as depicted in figure 1, with a maximum power factor of 0.9 as required by German national regulations [1]. The malfunction assumed in this scenario is a single PV inverter inverting the control curve parameterized due to a possible malfunction during a firmware update. This altered reactive power dispatch influences the voltage at the terminal the PV is connected to. This means the device is failing to inject reactive power in times of high active power dispatch meaning not exercising its grid support function. From a certain point in time on, the voltage is therefore not controlled accordingly. The impact left by this on the opertional data, such as the voltage, could then be used to classify in regular or erroneous behaviour.

The grid simulation software DIGSILENT Powerfactory [1] will be used to run grid simulations. These simulations can be of arbitrary length, weekly, monthly or an entire year. Grid simulation is the mean of choice to obtain data of the scenario described above, since such events are rare and are mostly not detected at their time of occurence, but at a later point in time. Simulation can help bolster the dataset especially with instances of data of cases in which a malfunction occured. For every simulation run, the PV inverter device which is experiencing a malfunction as well as when this malfunction occurs are chosen randomly. The scenario setup, as well as labelling the data as 'regular' or 'malfunction' will

[1] https://www.digsilent.de/de/powerfactory.html

be handled through a Python API. This allows for semi-automatic generation and labelling of data. A publicly available low voltage grid model will be used for generating the data in order to be able to later make the collected dataset publicly available. Such grid models are provided by the Simbench[2] project of the Uni Kassel. A grid such as the '1-LV-semiurb4–1-sw' with 45 busses can be employed, allowing for very life like simulation circumstances.

Data of this particular problem is interesting for the application of deep learning since voltage curves are highly non-linear and features are not easily deducted from these curves. Furthermore, the biggest variability in the data occurs during the usual functioning of the system, as for PVs in regular operation or households. Therefore, classification of unusual behavior might not be possible on the low voltage level with algorithms such as Support Vector Machine (pSVM)[8]. For this purpose, an artificial neural network (ANN) could be used, as described in [2]. Here, only Fast Fourier transformed voltage waveforms are required to train the ANN with normal and abnormal data. In this case faults of the inverter are targeted, not anomalous behaviour of the same. Using deep learning the raw voltage data could be used to train a network, learning the distinct features of the voltage curve at a terminal to which a PV experiencing a malfunction is connected. This pre-trained network could then detect malfunctions also in other scenarios as other grid setups.

2 Dataset

Data is to be acquired to develop means to detect and distinguish anomalies as wrongly parameterised PV inverters. Using the grid simulation setup descibed above allows for the collection and annotation of several hundred samples of timeseries data of households with PVs in a 15 minute resolution. This is a typical sample time for smart meters, that also allows using the data for classification of operational states [4]. The samples consist of timeseries data of voltages in 15 minute resolution. A sample could be a timeseries of one year (365 days), meaning every sample would have 10950 datapoints. Assuming 1000 samples collected this would yield a dataset with 10950000 numeric instances (the voltage values) and 1000 categoric values (the goal) of 'regular' or 'malfunction'. The simulation would obviously leave no misssing values in the dataset. As the samples are annotated with either 'regular behaviour' or 'malfunction', a sufficient amount of these samples should be of the class 'malfunction' to allow the extraction of features. 'Malfunction' is naturally the minority class, nevertheless about one quarter of the cases should be of this class on order to enable proper learning of features and classification [3]. Moreover, to mimic actual smart meter measurements, noise is added to the data created. The smart meters assumed, at which the data is collected, are standard smart meters with an accuracy of 1%, meaning the error on the reading is at max 1% of the smart meters highest displayable value. Therefore, a Gaussian white noise, with a

[2] https://simbench.de/de/datensaetze/

mean of 0 and a standard deviation such that the distributions biggest value is about the maximum error of the smart meter, is added to the data.

3 Schedule

Table 1 gives an overview of the workload projected for dataset collection, designing and building an appropriate network (ANN design), training and fine-tuning the network (ANN tuning), building an application to present the results (Results), writing the final report (Report) and preparing the presentation of the work (Presentation). A significant part of the time invested is projected for the design and tuning of the network. This is due to the higher uncertainty the carrying out of these tasks bears in comparison to the generation of the data.

Table 1: Overview of workload projected

Task	Dataset collection	ANN design	ANN tuning	Results	Report	Presentation
Hours	15	7.5	15	7.5	10	4

References

[1] Erzeugungsanlagen am niederspannungsnetz (vde-ar-n 4105) (2018)

[2] Khomfoi, S., Tolbert, L.M.: Fault diagnostic system for a multilevel inverter using a neural network. IEEE Transactions on Power Electronics **22**(3), 1062–1069 (2007)

[3] Krawczyk, B.: Learning from imbalanced data: open challenges and future directions. Progress in Artificial Intelligence **5**(4) (2016)

[4] Lu, N., et al.: Smart meter data analysis. In: PES T D 2012, pp. 1–6 (2012)

[5] Mahmud, N., Zahedi, A.: Review of control strategies for voltage regulation of the smart distribution network with high penetration of renewable distributed generation. Renew. Sustain. Energy Rev. **64**, 582 – 595 (2016)

[6] Vergara Barrios, et al.: Feasibility and performance assessment of commercial pv inverters operating with droop control for providing voltage supports services. In: 2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conf. (ISGT) Europe. IEEE Press (2019)

[7] von Appen, J., et al.: Time in the sun: The challenge of high pv penetration in the german electric grid. IEEE Power and Energy Magazine **11**(2), 55–64 (2013)

[8] Zhou, Y., et al.: Veto-consensus multiple kernel learning. In: Thirtieth AAAI Conf. on Artificial Intelligence (2016)