FRI Data Science Project Competition 2024
Interim report

Siemens Hybrid AI project

Adam Rychert, Urh Prosenc, and Evgenii Posashkov

Abstract

This overview dives into explanation of the decision-making processes of an electric grid system managed by Siemens. Optimization of energy usage is crucial in making the grid more sustainable and efficient. We perform an initial dataset exploration, revealing discrepancies between the dataset and provided information about the system. We establish a baseline model on a subset of data, successfuly recovering ground truth rules, governing the grids behavior. We highlight potential importance of solar power and propose further refinement and exploration to enhance model performance and uncover additional insights crucial for system optimization.

Keywords
Hybrid AI, Logic Tensor Networks, Neural Networks

Advisors: prof. Jure Demšar

Introduction

The motivation for optimizing energy usage lies in the need for sustainability, efficiency, and reliability in the energy infrastructure to mitigate climate change and ensure a brighter, greener future. The focus of this project relates to a power grid, which connects electric vehicle charging stations to three different power sources – solar panels, batteries and the public electric grid. Due to limitations of the grid and variable nature of solar power, a system was put in place by Siemens that governs the behavior of the grid and distributes power according to momentary supply and demand (see Figure 1). However, the almost black box nature of the decision system makes it difficult to further improve and optimize its decision making process.

Our goal therefore, is to explain the decision-making process that governs the behavior of the grid. We approach the task by first trying to recover the governing rules from the dataset provided by Siemens. Next, we look to further expand upon those rules. Finally, we look for ways to optimize the behavior of the grid via threshold optimization.

Data Exploration

The dataset, provided to us, came in the form of 1 month worth of hourly data, related to the electric grid. It contains information about power demand of 8 different charging stations, 1 of which supports fast charging, and power supply from the three sources - Solar panels, batteries and public electric grid. Additionally, battery state of charge is also provided. In the first steps of the exploration, dataset was examined in the light of the following ground truth rules, that were provided by Siemens that were provided by Siemens:

- If **Battery SOC** $>$ **80%**, then E-cars charging is completely covered by the local battery.
- If **40%** $<$ **SOC%**, then if **grid power** $>$ **limit**, e-cars charging power is covered by local battery.
- If **Battery SOC** $<$ **40%** and:

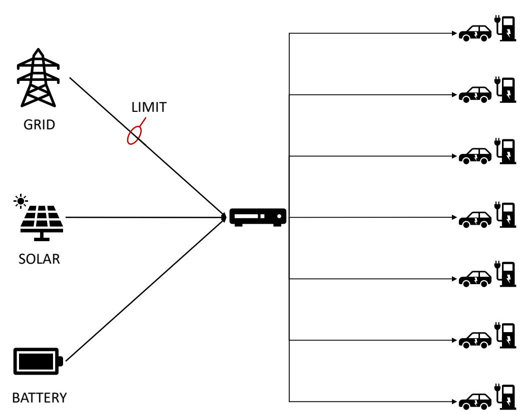

Figure 1. A visualization of the electric grid. A car using the charger gets energy from one of three sources – the grid, the solar panel, or the battery. Our goal is to uncover rules that decide which charging regime gets used depending on the condition of the system.

- **grid power** $>$ **limit**, then e-cars charging power is covered by local battery.
- **grid power** $<$ **limit**, then local battery is charged from the grid.

• If **Battery SOC** $<$ **15%**, then battery discharging is stopped due to battery health.

Several drawbacks of the limited dataset in relation to the ground truth rules were discovered. Firstly, there are no instances, where battery SOC would exceed 80% (probably because charging the battery above that is not good for its health), which makes the first rule impossible to recover. There are also several instances where despite power drawn from the grid exceeded the limit, the batteries were not fully utilized. Similarly, there are instances, where despite battery SOC $<$ 15%, the batteries are still providing power to the grid. In total, only approximately 16% of the data points were in line with the suggested ground truth.

Baseline Models

With those shortcomings in mind, we first decided to clean up the dataset and try to train a suitable baseline model on the subset of data, which strictly follows the provided ground truth rules. Decision trees were selected as the most appropriate model, mainly because of their simplicity and explainability. The nature of the model allowed us to extract exact rules and thresholds, governing the dataset. While classification accuracy is not the most important for the purpose of this project, it still provides a valuable insight into how well the model managed to capture the relationships in the data. After weighing the samples to account for class imbalance a decision tree of depth 2 was trained and tested on a random 80/20 train-test split of the dataset. The classifier, predicting the source of power on the cleaned dataset based only on battery SOC, achieved an accuracy of 71%, while the accuracy of the final model was 98%. The resulting model can be seen in Figure 2

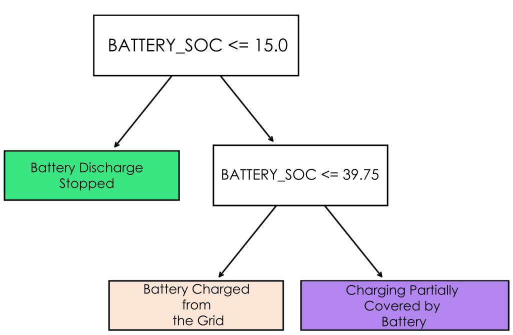

Figure 2. Baseline decision tree model, trained on the cleaned dataset. Both the 15% and the 40% tresholds are retrived.

While we were able to extract the lower threshold of $SOC = 15\%$, the upper bound came a bit lower than expected at $SOC = 39.75\%$, however, considering the nature of the algorithm and the size of the available dataset, that is not an issue. More importantly, only battery SOC was used as a feature when training the model. This was done with the purpose of extracting the ground truth rules from the model. However, feature performance, performed on the entire dataset, revealed that solar power also plays an important role. This was further comfirmed by two other separate models: XGBoost and Random Forest (see Figure 3). Despite its importance, solar power is not mentioned in the provided ground truths.

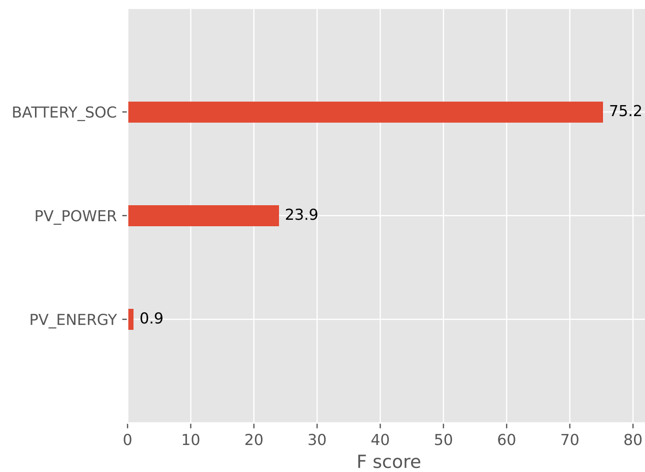

Figure 3. Feature importance based on gini index, extracted from XGBoost model. As expected, the battery SOC turned out to be the most important, followed by the solar power (with as much as 23.9%).

Further Work

While the initial exploration provided a valuable insight into the problem, there are several questions that still need to be answered, such as what part does the solar power play in the decision making process?

In the next step, our main focus will be on improving the baseline model. Further work needs to be done, including implementation of a neuro-symbolic model, such as LTN [1]. Possibilities of threshold optimization could also be explored.

References

[1] Luciano Serafini and Artur d'Avila Garcez. Logic tensor networks: Deep learning and logical reasoning from data and knowledge. *arXiv preprint arXiv:2012.13635*, 2016.