AI based Energy transition: data and model collection

Hou Shengren
hshengren@tudelft.nl
Intelligent Electrical Power Grids
Building 36, Room LB03.170

2022

1 Motivation

The energy system is undergoing a transition, from a centralized, fuel-based, automatic control-dominated system, to a decentralized, high renewable energy penetration, and autonomous entity. This transition requires advanced artificial intelligence (AI) technologies to reformulate the operation framework and simulation models, and fully dig the information behind the multi-dimensional, historical, operational data inside energy systems. These data contain hidden paradigms and valuable information, which was ignored before because of various reasons, like shortage of computing power and insufficient ability of the representative model.

AI-based energy transition highly depends on a large amount of high-quality data. Such as information about existing power stations, capacity, yearly electricity consumption, and ancillary service requirements, but also (hourly) time series of load, wind, and solar power generation, heat demand, etc. However, data collection is tedious. The bits and pieces of data are sometimes hard to find, often poorly documented, and almost always tedious to process: files are provided in different formats; downloading requires repetitive manual clicking; data structures between different sources are incompatible; daylight savings time and leap years are treated differently; URLs change frequently, and older data are updated without informing users (and sometimes deleted altogether).

Double work is inefficient. Currently, dozens, if not hundreds, of modeling teams in Europe spend significant resources gathering and processing data, all doing essentially the same thing. Highly skilled people waste a lot of time gathering data, time that would be better spent doing actual research. We have gone through this process ourselves, a sometimes quite frustrating experience. Double work is a waste of resources. Providing aggregated and aligned data in a central place is a public good that had been short in supply.

Moreover, data and models should be integrated together with a standard framework. That is to say, the entity should be stored in this way: one dataset is used to train one model with a specific task. Thus, the data-based model can directly run with data and dig information, to avoid a mismatch between the two parts. This match and "directly use feature" is vital for individuals from different backgrounds to get familiar with, learn, and explore the magic of data-based models in energy systems.

With inspiration from existing platforms, we create and develop this project, with the primary goal, of facilitating students interested in AI-based energy transition, to study and investigate this research area. The scope of data and models are limited to these inside EU.

2 Progress record

In this section, we record the progress of this project, which is updated every two weeks. Our goals are listed below:

1. Collect and explore existing data and model platforms and show features
2. Store this information in a meta-data structure.

2.1 Summary for collecting open-sourced platforms data and models

At this stage, the existing open-sourced data and model platforms are collected and investigated. We also explored and listed how these platforms stored data, models, and other features. The information is stored in a meta-data structure, here is JSON format.

2.1.1 concept for power system data

The databases themselves may furnish information on national power plant fleets, renewable generation assets, transmission networks, time series for electricity loads, dispatch, spot prices, cross-border trades, weather information, and similar.

They may also offer other energy statistics including fossil fuel imports and exports, gas, oil, and coal prices, emissions certificate prices, and information on energy efficiency costs and benefits.

With the integration of AI technologies and energy transition, more high-quality, structured, high-resolution data are urgently needed.

We explored the current data platforms and give a brief review below. Existing platforms have already collected a large amount of power system data, regarding **Electricity consumption** from hourly, to high resolution, like 1 min, **Renewable generation**, **market data**, including time-series spot prices, forward contracts, **technical data**, including installed capacity, the profile of power plants, specific network topology, and related parameters.

2.1.2 Concept for Power system model

Power system models invariably have a temporal resolution of one hour or less. Some models concentrate on the engineering characteristics of the system, including a good representation of high-voltage transmission networks and AC power flow.

Others models depict electricity spot markets and are known as dispatch models. While other models embed autonomous agents to capture, for instance, bidding decisions using techniques from bounded rationality. The ability to handle variable renewable energy, transmission systems and grid storage are becoming important considerations.

Nowadays, with the development of AI technologies, data-driven-based power system models are more and more popular. AI technologies have been implemented to solve various tasks in power systems, regarding optimal power flow, economic dispatch, energy management, false data injection, electricity forecasting, etc.

2.1.3 Problem of existing data and model platforms

After collecting and investigating current platforms of data and models. we see several drawbacks exist below.

- **Mismatch between data and models.** Data and existing models are hard to integrate. for example, currently, time-series electricity consumption and renewable generation are collected at either the national or residential level. However, the power system is a network, with various topologies. collected data ignored the network topology can not be directly used for a specific model.

- **Unclear guidance and documentation**, For most data and models, there is no clear documents and tutorial for organizing data, the interaction between data and model, how to use the model, etc.

In summary, existing data and model platforms that separate these two parts, are not suitable for AI-based energy transition. The AI-based energy transition requires tense-linked data and models because the performance of the model depends on the quality of data, and the interaction between data and model is frequent. A new platform integrating data and models in a suitable framework is vital for the progress of releasing the enormous potential of AI technologies in the energy transition.