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 should first feed the original data into the model, then preprocess the data, then update the parameters based on the data. This is called the "directly use feature", which 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.**

3 Case Study

In this section, we selected an important task of the power system research, i.e. congestion management to illustrate the framework and details for data and models used in the model-based based research and data-driven-based methods.

3.1 Congestion Management

The increased load demand and high penetration of renewable energies make distribution networks run on over-dimensioning conditions, leading to serious congestion problems, Congestion management thus is critical to enforce a safe, cost-effective, and sustainable power system operation.

Congestion problems in distribution networks can be envisaged as voltage problems (voltage should be limited in the safe boundary) and overloading problems (loading is close to or exceeds the thermal limits of power components). Modern distribution networks are weak networks with long feeders, where voltage problems are more common than thermal limits.

Thus, we use voltage regulation as a case of congestion management. Delivering active and reactive power through a feeder will lead to a voltage drop while penetration of local renewable energies will cause the over-voltage problem. The control methods for voltage regulation in distribution networks consist of network reconfiguration, reactive power control, and active power control. Network reconfiguration refers to changing the distribution network structure by adjusting the status of the switches, to deliver the power to customers in an efficient and regulated voltage condition. **Reactive and active power control methods refer to changing power flow to meet voltage constraints through demand response, power generation curtailment, or energy storage systems.**

Depicting data-driven and model-based approaches to voltage problems is vital for investigating the evolutionary trends of current research under AI-based energy transition background. In this section, we choose to compare the details of papers that solve voltage regulation problems with a model-based approach, and a data-driven approach (supervise learning). The commonalities and differences between the two approaches are summarized, which are illustrated based on model, data, and algorithm categories. The bounds between the model, data, and algorithm are identified below.

- The model is a surrogate of the realistic entity, which should be created by accurate parameters and mathematical formulation. In voltage regulation problems, the model is an electricity network with a specific topology and parameters. parameters used for creating a model can also be seen as a kind of data.

- Data refer to the information that can be measured and collected explicitly and is used to record the model condition during the electricity generation, transmission, and load consumption dynamics. In voltage regulation problems, input data include reactive and active load consumption of each node, while output data is the voltage of each node, which is calculated by solvers executing power flow estimation.

- Algorithm is used to determine the values of variables with a specific goal, given the input data and model. In voltage regulation problem, the decision variables can be power curtailment of PV generation, charge/discharge of storage system, and the goal is to limit the voltage into predefined boundaries like $0.95pu - 1.05pu$

Note: no matter whether using model-based or data-driven based approaches, the goal is to develop a new algorithm (method) that can meet the voltage regulation demands in a realistic electricity network, i.e. real-time response, optimal solution, etc.

3.1.1 Case study for Model-Based Voltage Regulation

In the research[1], smart (photovoltaic) PV inverters are used to control power flow to meet the voltage constraints by adjusting the active power output from PV inverters in an active distribution network. First, the model of the active distribution network and energy units in this network is built based on parameters and mathematical formulations. It includes the parameters of line, distribution transformer, PV inverters. Then one day's load consumption data and PV generation data on each node are input into the model to show the voltage violation situation without control of active power.

As the distribution network is a radial network, the author finds that the voltage drop problem of one node can be solved optimally by decreasing the active power of the neighboring node with low voltage. For voltage rise problems, adjusting the node itself or the neighbor nodes have similar performance.

Based on this knowledge and analysis of the network, a method is proposed that can solve the voltage regulation problem in a decentralized and real-time way. During the research, **the input data is not vital and follows a one-time use principle because the method developed just depends on the knowledge and analysis of the model instead of the data**. **Finding out the optimal solution is based on the model.**This is the main reason why it is called a model-based approach. The method is highly linked to the model because it needs to estimate the model dynamics and then find out the optimal values of decision variables during the online-schedule stage.

3.1.2 Case Study for Supervise Learning Based Voltage Regulation

A large amount of labeled data including is required for supervised learning-based voltage regulation problems[2]. The essence of these methods is to build a mapping between the input (load consumption and PV generation of each node) to the output (the optimal value of active power control), which is fully determined by the weights and parameters of the surrogate function (linear function or deep neural networks). In order to improve the performance and scalability, more data, representing different scenarios is required to update the weights and bias of the deep neural networks. The label of data is the optimal value of decision variables, which can be gathered by iterative solving the model-based optimization approaches.

During the online schedule stage, the trained deep neural network can directly get the optimal value of decision variables without building a model and solving it as the last case did. Because the deep neural network has built a black box function based on the combination of weights, and bias, representing the model dynamics and correspondence between voltage and the decision variables, and the optimal goal.

Thus, the supervised learning-based voltage regulation algorithm is highly linked with data instead of the model. It uses labeled data to train parameters of deep neural networks, indirectly building a surrogate model while model-based approaches through accurate parameters and mathematical formulations.

3.1.3 Discussion

The general procedure of solving voltage regulation problems by using model-based approaches and data-driven is illustrated. We divided such research into three components: model, data, and algorithms, with a clear boundary and show the commonalities and differences between supervised learning and model-based research cases. Model-based algorithms are highly dependent on *model* because the performance of the algorithm is based on the accuracy of the model. For data-driven methods, the performance of algorithms depends on the quality of data, even when the model is still needed when preparing the label. Nowadays, researchers are more and more focusing on data-driven methods to prompt the AI-based energy transition. But this trend can be attributed to two root reasons: 1. **a large amount of data has been accumulated in energy systems** 2. These data now can be utilized to get useful information and help make decisions, based on the flourishing development of machine learning technologies, especially deep learning.

References

[1] A. Fu, M. Cvetković, and P. Palensky, "Distributed cooperation for voltage regulation in future distribution networks," *IEEE Transactions on Smart Grid*, 2022.

[2] P. Vongvilasack, S. Premrudeepreechacharn, and K. Ngamsanroaj, "The application of machine learning for the voltage and reactive power control in power distribution network," in *2022 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 2022, pp. 1–6.