

Neural network to generate synthetic building electrical load profiles

Energiesystem- und Klimamodellierung

Francesca CONSELVAN⁽¹⁾, Philipp, MASCHERBAUER⁽²⁾, Daniel, HARRINGER⁽²⁾

⁽¹⁾ e-think energy research, ⁽²⁾ EEG, TU WIEN

Motivation und zentrale Fragestellung

At our days, we can collect a considerable amount of data in the energy sector, but, due to privacy concerns, companies are unable to share customer meter data for general research and analysis. Synthetic data provides a valid alternative to real data since they anonymise the data and maintain its original statistical information. This paper presents a real case scenario of the generation of circa 50 000 synthetic energy load profiles, addressing the following questions:

1. What are the exogenous variables of the load profiles, like building and vintage type that can be used to generate synthetic data?
2. What is the best fitting methodology to generate synthetic data for energy load profiles?
3. What are the limitations and the challenges in using the chosen methodology?

This research is carried out in the framework of the Horizon Europe project MODERATE.

Methodische Vorgangsweise

In the preliminary results, we have analysed circa 50 000 annual electric load profiles, following the below methodology:

1. Data Standardisation. Each load profile is normalized by its annual peak load to make sure that different buildings are similar in magnitude.
2. Data Clustering. Load profiles vary a lot, even for the same building. So, load profiles are clustered to capture the key variables and characteristics. Different clustering methods are used and compared, specifically:
 - Hierarchical based: hierarchical clustering groups
 - Centroid base: k-means and agglomerative cluster, whose optimal value of clusters is determined with Elbow and Silhouette methods
 - Principal Component Analysis (PCA)
 - Density-based: DBSCAN and HDBSCAN
3. Synthetic data generation. We generate the synthetic data using the neural network algorithm GANs (Generative Adversarial Networks), which consists of two neural networks, a generator and a discriminator, contest with each other.

Ergebnisse und Schlussfolgerungen

The first step is to select the optimal number of clusters, which is 10 for k-means and 5 for PCA. Some clusters demonstrate a clear working day pattern with energy usage increase in the morning and decrease in the late afternoon.

The second step is to evaluate the accuracy of the generator and the discriminator, which we expect to increase with the increase of the training epochs. We expect that at the beginning of the training the accuracy of the discriminator is high, and it decreases with the training process going on, because the generator can generate load profiles that are more like the real ones. We aim to determine the number of the epochs that benchmark the accuracy of the generator and the challenge for the discriminator to discern fake from real load profiles.

Literatur

- [1] Wang, Zhe, and Tianzhen Hong. 'Generating Realistic Building Electrical Load Profiles through the Generative Adversarial Network (GAN)'. *Energy and Buildings* 224 (October 2020): 110299. <https://doi.org/10.1016/j.enbuild.2020.110299>.

- [2] Figueira, Alvaro, and Bruno Vaz. 'Survey on Synthetic Data Generation, Evaluation Methods and GANs'. *Mathematics* 10, no. 15 (2 August 2022): 2733.
<https://doi.org/10.3390/math10152733>.
- [3] Behm, Christian, Lars Nolting, and Aaron Praktiknjo. 'How to Model European Electricity Load Profiles Using Artificial Neural Networks'. *Applied Energy* 277 (November 2020): 115564.
<https://doi.org/10.1016/j.apenergy.2020.115564>.
- [4] Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 'Generative Adversarial Nets', n.d., 9.