Applying transformer to imputation of multivariate energy time series data

Hasan Ümitcan Yilmaz [1] **Max Kleinebrahm** [1] **Christopher Bülte** [1] **Juan Gómez Romero** [2]

Abstract

To reduce the greenhouse gas emissions from electricity production, it is necessary to switch to an energy system based on renewable energy sources (RES). However, intermittent electricity generation from RES poses challenges for energy systems. The primary input for data-driven solutions is data on electricity generation from RES, which usually contain many missing values. This proposal studies the use of attention-based algorithms to impute missing values of electricity production, electricity demand and electricity prices. Since attention mechanisms allow us to take into account dependencies between time series across multiple dimensions efficiently, our approach goes beyond classic statistical methods and incorporates many related variables, such as electricity price, demand and production by other sources. Our preliminary results show that while transformers can come at higher computational costs, they are more precise than classical imputation methods.

1. Introduction

According to the latest breakdown of global emissions published by Climate Watch and the World Resources Institute, the electricity sector accounts for 24% of global emissions, while the energy sector accounts for 73% (Our World in Data, 2021; Climate Watch, 2021; World Resources Institute, 2020). Undoubtedly, electricity production by renewable energy sources (RES) plays a key role to mitigate climate crisis. Furthermore, RES can help to reduce emissions in other sectors through sector coupling, such as using electric cars in transportation or electric boilers and heat pumps in heat sector. Data on electricity generation from RES is a key element addressing climate change. RES generation data are used in many contexts, ranging from studies of technical and economic feasibility of the expansion of the RES to short-term prediction of energy production and optimal control or efficient use of this electricity in various sectors. Therefore, improving the quality of these data ensure greater accuracy and consistency of research in this area, and will open new avenues for future innovations.

Several electricity time series have been made available to the public. Two important examples are the PJM data (PJM data miner, 2021) in the US and the transparency platform ENTSO-E (ENTSO-E, 2021) in Europe. However, compiling these data for a very large number of electricity production units, including small generation such as rooftop solar plants, is difficult and produces time series with many missing values. For instance, in the ENTSO-E transparency platform an average of 1000 values per week are missing in the production mix data for 2015 and 2016 (Hirth et al., 2018).

(Ruggles et al., 2020) show that after filling the gaps only in electricity demand time series, the results of a power system model vary by 5% between using two sophisticated data imputation approaches, even for a very simple analysis considering only one region. They also show that the results from simple data imputation methods are very high and unacceptable. In this work, we fill the data gaps in the data of the ENTSO-E transparency platform, which consists of 32 regions (European countries), and we fill the gaps not only in electricity demand but also in time series of electricity generation such as wind and solar and electricity prices.

Time series literature have used different methods for data imputation in the past, mostly based on statistical correlation and regression techniques (Van Buuren, 2018). More recently, the advances in Machine and Deep Learning suggest that sequence prediction and forecasting methods can be successfully applied to imputation of missing values in energy data (Wang et al., 2019). In particular, Recurrent Neural Networks (RNNs) and Long-Short Term Memory (LSTMs) networks have been applied both to forecast electricity production from renewable energy sources and electricity demand, and to fill in missing values (Liu et al., 2019), (Kumar et al., 2018).

In the last years, transformers (Vaswani et al., 2017) have become the choice architecture to process sequences in nat-

[1]Institute for Industrial Production (IIP), Chair of Energy Economics, Karlsruhe Instituteof Technology (KIT), Germany [2]Department of Computer Science and Artificial Intelligence, University of Granada, Spain. Correspondence to: Hasan Ümitcan Yilmaz <hasan.yilmaz@kit.edu>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

ural language processing applications. Transformers use an attention mechanism to identify the relevant fragments in a sequence. In this way, transformers learn dependencies between all considered time stamps explicitly.

In the case of an electricity system, there are high dependencies between the electricity generation, electricity demand and electricity price time series. As an example, the following interdependencies can be mentioned. In a power system, the dispatch of the different power generation technologies depends on the dispatch of the competing technologies. The sum of electricity generation must always match electricity demand, taking into account storage technologies. In addition, electricity prices in each hour depend on the dispatch of the electricity generation technologies in that hour. Therefore, there are sophisticated interdependencies between time series that can be understood through an innovative attention mechanism transformers. The attention mechanism makes it possible to learn direct dependencies between individual features at certain timestamps while RNN-based approaches use a compressed mapping of the time series as the basis for the imputation. By using direct dependencies instead of a compressed representation of the timeseries transformers can be trained faster and achieve better metrics (Sucholutsky et al., 2019).

In contrast to the literature, we use all data on electricity generation, electricity prices and electricity demand to fill the gaps in any of these time series. Furthermore, we analyze the problem as a whole and try to fill the gaps in the data for all power generation technologies, electricity demand and electricity prices. This enables us to fill gaps across all data dimensions simultaneously. An overview of the architecture of our time series imputation transformer model can be seen in Figure 3.

2. Data imputation for energy time series

Year	Country	Hour	S_1	S_2	...	S_n
	C_1	H_1	.	.	.	.
		H_2	.	.	.	.
		...	...	...	...	...
		H_{8760}	.	.	.	.
	C_2	H_1	.	.	.	.
		H_2	.	.	.	.
		...	...	...	...	...
Y_1		H_{8760}	.	.	.	.

Figure 1. Demonstration of the input data

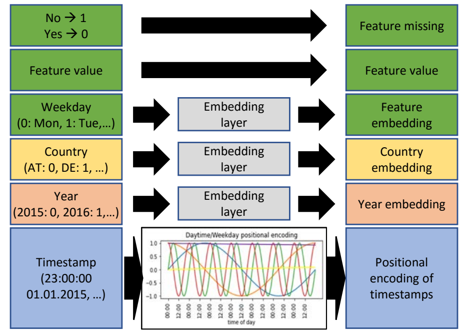

Figure 2. Visualization of the first input processing layers

We use the inter-dependencies between the electricity prices, electricity demand and production mix for data imputation. A detailed analyses of the impact of renewable energy sources on the price levels and variability can be found in (Maciejowska, 2020). The table in Figure 1 shows the structure of the input data. Let $C = \{C_1, \ldots, C_n\}$ be a set of n countries (e.g., "France", "Germany") and $S = \{S_1, \ldots, S_n\}$ a set a n time series types (e.g., "electricity production from solar", "electricity prices"). $Y = \{Y_1, \ldots, Y_n\}$ represents the years and $H = \{H_1, \ldots, H_{8760}\}$ the hours of the years. We are working with 21 different electricity generation technologies, electricity demand and electricity prices as input time series S.

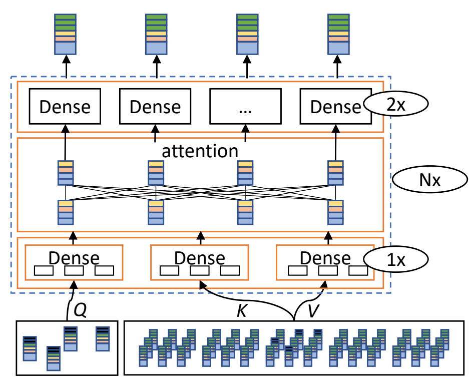

Figure 3. Main architecture of the transformer model

Figure 2, demonstrates the processing of these time series before it is provided as input to the transformer layers. Information about each value at the specific time step as well as information about the availability of that feature is directly fed into the concatenation layer. Categorical information about the feature number, country, year and weekday is first inserted into embedding layers which learn multidimensional representations of the input during training. In addition, the model is supplied with three different positional encodings to make it easier for the model to capture daily, weekly and seasonal rhythms. This is especially helpful since the time series under consideration can have strong seasonal characteristics, e.g. Solar or Wind.

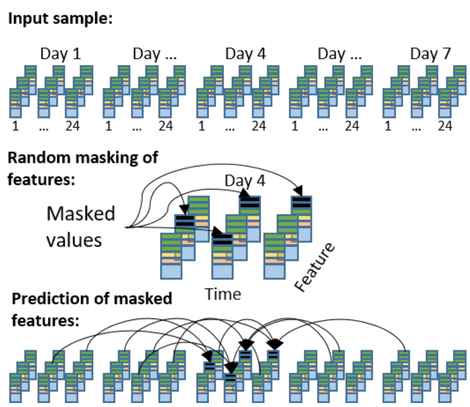

Figure 4. Exemplary visualization of the experimental setup of the transformer model

During the training process of the imputation model, we input samples which cover a period of one week in hourly time resolution. Each timestep contains 22 features. As a result, there are 3,696 data entries in each sample, creating a self-attention map of size of 3,696 x 3,696. During the training process we randomly mask x data entries of the middle day of the week (x ¡= 22 x 24). These masked values are then predicted by the model. This approach can be found in Figure 4. It is worth noting that such a model requires a lot of memory during training. Therefore, it is planned to integrate an attention decomposition mechanism, that reduces the model size (see for example Ma et al. (Ma et al., 2019)).

3. Preliminary results

To determine the performance of our approach, we use three typical imputation methods as a comparison. The first one is called Last Observation Carried Forward (LOCF), which takes the last x values as a prediction, in our case, the last 24 hours. The second method is ridge regression, which is a specialized form of linear regression. Finally, the third comparison method is the so called KNN-Imputer. This is a autoregressive method, that measures the similarity between each day of the time series to find similar days as a prediction. The latter approach has shown to be a popular method for data imputing (Kuhn et al., 2013).

For these preliminary results, we only worked with a subset of the data from ENTSO-E, that only includes data for Germany from 2018-2020. Since small gaps in the data can be easily recovered using simple imputation methods, e.g linear interpolation, we want to focus on larger gaps. This is based on our finding that 86.7% of missing values in the ENTSO-E data from 2015-2020 are made up of gaps equal or larger than 24 hours. To get an overview of the possibilities of the transformer model, we first study its ability to impute gaps with the length of 24 hours. Therefore, we create samples with a timespan of one week by always shifting the week one day to the right. We then remove information of the feature under consideration (electricity production from solar) of the day in the middle of the week.

An example of how the imputed values compare to the original can be found in Figure 5. The error in the preliminary results using mean squared error (MSE) with Transformer are the lowest (0.0585). After that, KNN-Imputer performs the best (0.0629). The errors of LOCF (0.1133) and Ridge Regression (0.1827) are significantly higher. In addition, an important advantage of the transformer is allowing simultaneous imputation of values across different dimensions which is very useful when data for multiple features are missing. However, the transformer can incur very high computational costs.

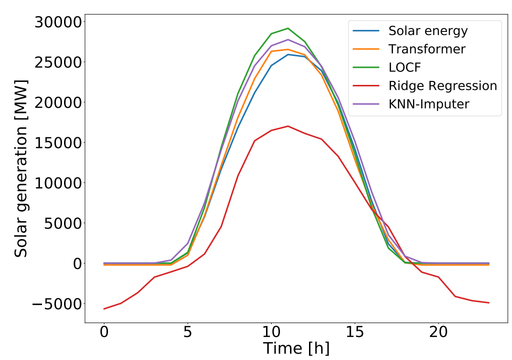

Figure 5. An example of the prediction of the feature Solar

4. Summary and outlook

We successfully used an attention based transformer algorithm to fill the data gaps on electricity generation from renewable energy sources considering cross-dimensional dependencies in different types of electricity data. Improving the quality of these data is crucial as it is used in many fundamental analyses and energy models. Furthermore, this leads to an opportunity to overcome the barriers for the expansion and use of renewable energy sources and therefore has a direct impact on achieving climate goals. In our test case, we focus on Solar power generation in Germany, but this approach is easily expandable to Europe, or to different market regions, e.g. PJM. Since our results are quite promising, the next step is to evaluate the model on the complete data and to compare the results with other methods.

Acknowledgement

This work was supported by the OptTech Project (ID: A33419) founded by Stiftung Energieforschung Baden-Württemberg and J. Gomez-Romero was partially supported by the Spanish Ministry of Science, Innovation and Universities (TIN2017-91223- EXP).

References

Climate Watch. World total including lucf greenhouse gas (ghg) emissions climate watch. https://www.climatewatchdata.org/ghg-emissions?end%5C_year=2016%5C&end_year=2018&start%5C_year=1990&start_year=1990, 2021. (Accessed on 05/27/2021).

ENTSO-E. Entso-e transparency platform. https://transparency.entsoe.eu/, 2021. (Accessed on 05/27/2021).

Hirth, L., Mühlenpfordt, J., and Bulkeley, M. The entso-e transparency platform – a review of europe's most ambitious electricity data platform. *Applied Energy*, 225:1054–1067, 2018. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2018.04.048.

Kuhn, M., Johnson, K., et al. *Applied predictive modeling*, volume 26. Springer, 2013.

Kumar, S., Hussain, L., Banarjee, S., and Reza, M. Energy load forecasting using deep learning approach-lstm and gru in spark cluster. In *2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT)*, pp. 1–4, 2018. doi: 10.1109/EAIT.2018.8470406.

Liu, Y., Guan, L., Hou, C., Han, H., Liu, Z., Sun, Y., and Zheng, M. Wind power short-term prediction based on lstm and discrete wavelet transform. *Applied Sciences*, 9(6), 2019. ISSN 2076-3417. doi: 10.3390/app9061108. URL https://www.mdpi.com/2076-3417/9/6/1108.

Ma, J., Shou, Z., Zareian, A., Mansour, H., Vetro, A., and Chang, S.-F. Cdsa: cross-dimensional self-attention for multivariate, geo-tagged time series imputation. *arXiv preprint arXiv:1905.09904*, 2019.

Maciejowska, K. Assessing the impact of renewable energy sources on the electricity price level and variability–a quantile regression approach. *Energy Economics*, 85: 104532, 2020.

Our World in Data. Emissions by sector our world in data. https://ourworldindata.org/emissions-by-sector, 2021. (Accessed on 05/27/2021).

PJM data miner. Data miner 2. https://dataminer2.pjm.com/list, 2021. (Accessed on 05/27/2021).

Ruggles, T. H., Farnham, D. J., Tong, D., and Caldeira, K. Developing reliable hourly electricity demand data through screening and imputation. *Scientific data*, 7(1): 1–14, 2020.

Sucholutsky, I., Narayan, A., Schonlau, M., and Fischmeister, S. Pay attention and you won't lose it: a deep learning approach to sequence imputation. *PeerJ Computer Science*, 5:e210, 2019.

Van Buuren, S. *Flexible imputation of missing data*. CRC press, 2018.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. Attention is all you need. *arXiv preprint arXiv:1706.03762*, 2017.

Wang, H., Lei, Z., Zhang, X., Zhou, B., and Peng, J. A review of deep learning for renewable energy forecasting. *Energy Conversion and Management*, 198:111799, 2019. ISSN 0196-8904. doi: https://doi.org/10.1016/j.enconman.2019.111799.

World Resources Institute. 4 charts explain greenhouse gas emissions by countries and sectors — world resources institute. https://www.wri.org/insights/4-charts-explain-greenhouse-gas-emissions-countries-and-sectors, 2020. (Accessed on 05/27/2021).