
Exposé - Masterarbeit

Jan Besler
5629079

Abstract

With the advent of LLMs, the underlying architecture of transformation models is being used in the field of time series forecasting. One paper hypothesised that transformers work best with complex data such as languages and text data, while other models are advantageous for less complex data. This paper will explore this hypothesis using data of varying complexity and diverse models to determine which models work best with which data.

1 Research Problem

In early 2023, Large Language Models (LLM) attracted much public attention with the emergence of ChatGPT. The development of these models can be traced back to the development of transformer models, which have proven useful in the field of NLP. These transformer models use sequential data as input, i.e. a data point depends on the previous data points. As a popular model in NLP, transformer models are also popular in other fields, such as time series forecasting, where the data is sequential by definition. In this research area, other models still receive a lot of attention, for example Long Short Term Memory (LSTM). The question therefore arises why Transformer models are not widely used for time series forecasting? This question was investigated by Zeng et al., 2022 and their hypothesis is that it depends on the complexity of the underlying data which model performs better. They concluded that transformers only have an advantage when they work with complex data structures, such as languages and text data. Thus, a transformer model with less complex data may lead to worse results compared to another model such as an LSTM.

In order to investigate the hypothesis put forward by Zeng et al., 2022, this paper will test different models on data with different levels of complexity. The data chosen is from wind turbines for the more complex side and electricity consumption from different companies on the less complex side. These applications were chosen to investigate how predictive models can be improved. And whether it is generally worthwhile to develop more complex models for different use cases if they might give worse results than previous solutions.

In the area of electricity generation and consumption, forecasting is a relevant issue to drive the decarbonisation of the economy through the deployment of renewable energy. With the introduction of more renewable energy systems, electricity production fluctuates more than before, while demand patterns remain largely the same. This leads to conflicts, as demand and production need to be perfectly matched at all times. In addition, the stability of the power grid must be taken into account, as a production peak can lead to problems with the frequency of the power grid, which in turn can lead to system-wide failures and blackout outages. To avoid these problems, it is important to predict electricity production with high accuracy in order to know when the flow of electricity needs to be regulated by the relevant authorities and companies. On the other hand, it is beneficial for companies to know when they will consume the amount of energy for the corresponding time. This makes it easier to plan ahead, to take advantage of lower prices on the spot market and is another contribution to the stability of the electricity grid.

The deep learning models used are probabilistic and not deterministic. This is due to the advantages of probabilistic models when working with noisy data and the resulting probability distributions as

outcomes instead of deterministic values. With the ability to quantify the uncertainty of forecasts, there is an advantage to modelling in a probabilistic way by understanding the range and probability of possible outcomes. Even if there is no exact value for the outcome, ranges can be given to estimate how big the impact will be. However, the disadvantages of working with probabilistic models are the increased complexity due to working with probability distributions as outcomes, making the results less intuitive to interpret. In addition, the increased complexity leads to higher computational costs and thus greater demands on limited resources.

2 Data

Two data sets are used to compare the data complexity types. On the more complex side is SCADA wind farm data from Plumley, 2022, which is 10-minute SCADA and event data from the 6 Senvion MM92s at the Kelmarsh wind farm, grouped by year from 2016 to mid-2021. On the less complex side is electricity consumption data from 4 manufacturing companies in Germany provided by Fraunhofer IPA. The data is available in 15-minute intervals and shows the electricity consumption in a hierarchical representation.

3 Related Work

4 Research Question

The overarching goal is to investigate whether the transformer models bring such advantageous properties in terms of accuracy to justify the added complexity, as it has successfully happened in the NLP context. The aspects to focus on are the direct comparison with another deep learning model to evaluate if the transformer is actually performing better than a less complex LSTM model. And also how much time and resources the development takes up to come to a qualitative conclusion if a transformer model is worth the effort for the potential benefit versus other models.

Therefor a specific research question could be **"Are transformer models in time series forecasting worth the hussle? A data driven approach"**.

5 Methodology

The models selected for evaluation are several different transformer models, an LSTM model and a non-machine learning based model that serves as a benchmark for the other models. For the benchmark task, the Random Forest was chosen because of its numerous applications in the field of time series forecasting in previous work. The models that use deep neural networks are not written from scratch, but come from various publicly available libraries (PyTorch). These basic models are then fine-tuned using hyperparameter search. For the transformer models, various models from renowned papers as well as a self-made shallow transformer are selected to test whether the depth of the network has a large impact on the results. The forecast horizons are 1 hour, 24 hours and 96 hours, with the last forecast being the most difficult task. This last number was chosen because the weather forecasts are also produced for a 96-hour period. This rating is based on the total computational time to evaluate all models and the computational resources in terms of GPU/CPU and RAM usage.

It is difficult to quantify whether a model provides a benefit in terms of increased predictive accuracy compared to the computational costs spent on development. Due to scalability, a small increase in accuracy can justify huge development costs when deployed at scale. The data is first pre-processed in a data pipeline to detect outliers and find missing values that could later affect the models. In a second step, the data is fed into the different models, and each of the datasets is compared against certain metrics to measure the performance of the models. Finally, the results are printed in tabular form and presented graphically for better comparability in the paper.

References

- Plumley, C. (Feb. 2022). *Kelmarsh wind farm data*. Version 0.0.4. DOI: 10.5281/zenodo.7212475.
- Zeng, A., M. Chen, L. Zhang, and Q. Xu (2022). “Are transformers effective for time series forecasting?” In: *arXiv preprint arXiv:2205.13504*.