IFBD: Graph Convoluational Networks with Transformer for Long Sequence Predictions

Xiaotian Yu
yu.xiaotian@intellif.com

ABSTRACT

Wind Power Forecasting (WPF) is to predict a long sequence of power given historical records of wind turbines. We solve the problem of wind power forecasting by deep learning techniques in Baidu KDD Cup 2022, which last from March 16 to July 19 2022. In the challenge, we predict the effective power of each turbine for two days with a time interval of ten minutes. The problem of WPF requires models not only to learn the time dependency, but also to capture the spatial relations of all turbines in the wind farm. We propose a model based on graph convolutional networks (GCN) with transformer. Inspired by the latest Autoformer [19], our model learns the short-term and long-term dependencies. The proposed GCN-Transformer achieves a score of 46.65 at the final phase of Baidu KDD Cup 2022 under the framework of paddlepaddle. We also develop baseline models based on GRU unit. Finally, we achieve the score of 46.34 by the ensemble of three models.

1 INTRODUCTION

The task of Baidu KDD Cup 2022 is to predict the wind power supply of a wind farm. The wind farm consists of 134 turbines located at different positions. Turbines generate clean and safe renewable energy, but usually have a risk of variability in the output. The challenge of the Wind Power Forecasting (WPF) lies in an accurate analysis of grid stability and security of supply. Also, the spatial and temporal characteristics should be modelled to the prediction of the wind power.

In Baidu KDD Cup 2022, a new dataset called SDWPF [22] is released. There are 134 turbines in SDWPF with their spatial distribution in a farm. The effective power of each turbine, as well as the total power of the farm, is affected by dynamic context factors like temperature, weather, and turbine internal status. We note that, for the task of WPF, many existing datasets only contain time sequence information. Thus, the resultant methods learn the time dependency of WPF without the locations and context information of wind turbines. Intuitively, more information of wind turbines should provide important hints for the prediction of effective power. The SDWPF dataset provides the information about the wind, temperature, turbine angle and historical wind power. New methods could be developed based on the SDWPF dataset.

Baidu KDD Cup 2022, Mar. 16 - July 19, 2022

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

For the SDWPF dataset, the time range of is over 245 days. All samples are collected with a time interval of ten minutes. The spatial distribution of 134 turbines is also released.

The problem of WPF is technically long sequence prediction. According to the spatial scale of the wind power, the problem can be categorised as a single wind turbine, a wind farm and a group of wind farms. In this challenge, we aim to predict wind farm power for each turbine. Inspired by the technique of transformer in sequence prediction [19], we propose a model based on graph convolutional networks (GCN) with attention structures. The proposed model is named as GCN-Transformer. GCN-Transformer captures long-term and short-term dependencies, as well as the spatial relations of all turbines in a static graph.

In Baidu KDD Cup 2022, our GCN-Transformer achieves 46.65 at the final phase. By ensemble learning with baseline models, we finally achieve 46.34.

In this report, we have the following contributions.

- We reformulate the problem of long sequence prediction for wind power forecasting as a forecasting problem of graph learning with transformer. The number of turbines is 134.
- We propose a GCN-Autoformer network structure to capture the short-term and long-term characteristics among the wind farm turbines.
- We achieve good performance in the competition with the proposed method under the framework of paddlepaddle.

2 RELATED WORK

Predictions based on time series is essential in many practical applications. With historical data, predictions with time instants require models to learn different scales of time dependencies. The ubiquitous applications include power prediction, grid management, traffic flow prediction and portfolio management.

2.1 Long sequence predictions

Recently, sequence modeling is greatly powered by the technique of transformer [16]. It has the abilities of global attention and the powerful feature representation. Artificial intelligence methods with transformer have shown that machines could outperform human beings in many tasks, such as natural language processing [2], and speech recognition [3].

The survey of transformer-style methods has shown its great potential in predictions of time sequence [18]. For long sequence predictions, there are several interesting works, such as LogTrans [10], Informer [21], Autoformer [19], Pyraformer [12].

The problem of predictions for long sequence is significant. It has attracted various researchers and practitioners [10, 12, 19, 21]. For the problem of WPF, various models have been developed from viewpoint of spatial or temporal. The previous models can be classified into three categories: statistic models [14], machine

Table 1: Statistics of the dataset in Baidu KDD Cup.

days	# of used variables	# of records	ratio for training
245	10	4,727,520	86%

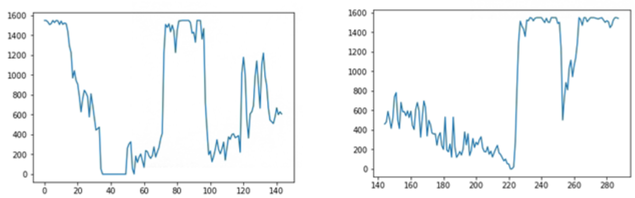

Figure 1: Plots of consecutive two days of power for Turbine 1. The left plot is the first day and the right plot is the second day.

learning methods [6] and deep learning methods [10, 21]. The most widely used model for long sequence prediction is autoregressive integrated moving average (ARIMA). Recently, many advanced models based on deep learning for long sequence prediction have great potential to tackle this problem, e.g., the transformer-style methods [12, 19].

In this challenge, we also fully take advantage of transformer structure for long sequence predictions.

2.2 Graph Neural Networks (GNNs)

In [15], Sperduti et al. first applied the structure of neural networks into learning patterns of complicated data, motivating the early studies of Graph Neural Networks (GNNs). The concept of GNN was formed in [4], and GNNs were firstly investigated for learning relation data with generalization ability in [13].

Recently, inspired by the success of deep convolutional networks in the computer vision domain, Kipf and Welling proposed a novel design of convolutional networks for graph-structured data [7]. Basically, deep convolutional networks effectively learns internal representation with spatial information [8]. For graph-structured data, it is essential to learn the feature representation of nodes and edges by propagating neighbor information [1, 9]. Later, efficient training of deep Graph Convolutional Networks (GCNs) has been widely studied, such as the model of GraphSAGE in inductive learning of graph [5]. By leveraging the self-attention layers, GNNs acquire a more powerful representation ability in dynamics of graph data [11, 17, 20].

In this report, we propose a graph convolutional network with transformer for wind power forecasting.

3 DATASET AND METRIC

The SDWPF data [22] are sampled every 10 minutes from each wind turbine in the wind farm which consists of 134 wind turbines. The statistics of the important information of the SDWPF dataset are shown in the Table 1.

3.1 Stability

We show the high variability of the sequence with the SDWPF dataset. The statistics of consecutive two days are shown in Figure 1. From the figure, we find that the task of wind power forecasting is challenging.

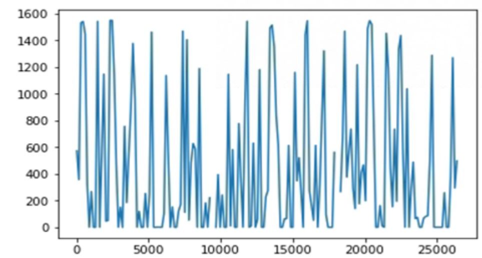

Figure 2: The long-term data characteristics of Turbine 1.

3.2 Long-term characteristics

By considering all the data over the whole time period, we show the long-term characteristics in Figure 2. We clearly find that there are some periodic signals for long-term information.

3.3 Metric

In this challenge, the metric is the average of MAE and MSE. During training the model, we set the loss as MSE.

4 METHOD

We develop a method of GCN-Transformer to predict the long sequence of wind power. Basically, the GCN-Transformer is inspired by Autoformer.

4.1 Overview of our method

The structure of our model is shown in Figure 3. We generate the predictions of all turbines in one model. The short-term relations are fed into the first transformer and the long-term relations are fed into the second transformer. We combine the two Transformers into an MLP structure.

Given the input feature as $x \in R^{134 \times p \times z \times q}$, where p is the input length, z is the batch size and q is the number of variable. In our model, we set $p = 144$, $z = 32$ and $q = 10$.

Then, with the framework of paddlepaddle, we construct graph convolutional network with attention. We have $x_{\text{att}} = GAT(x, G)$, where G is the graph under correlation coefficient. We decompose the sequence x into short-term and long-term terms. By applying the framework of Autoformer, we have $y = Transformer(x_{\text{att}})$.

4.2 Configurations of our method

The dimension of the hidden layer is set 128. The learning rate is set as 5×10^{-5}. The number of heads in transformer is set as 8. The number of layers for transformer in short-term relations is set as 2. The number of layers for transformer in long-term relations is set as 1.

5 EXPERIMENTS

We implement the model shown in Figure 3 under the framework of paddlepaddle. We have the following results.

The GRU and Graph-former are two released baselines in Baidu KDD Cup. Our model is GCN-Transformer and achieves 46.65 at the final phase. With ensemble learning on three models, we achieve the performance score of 46.34.

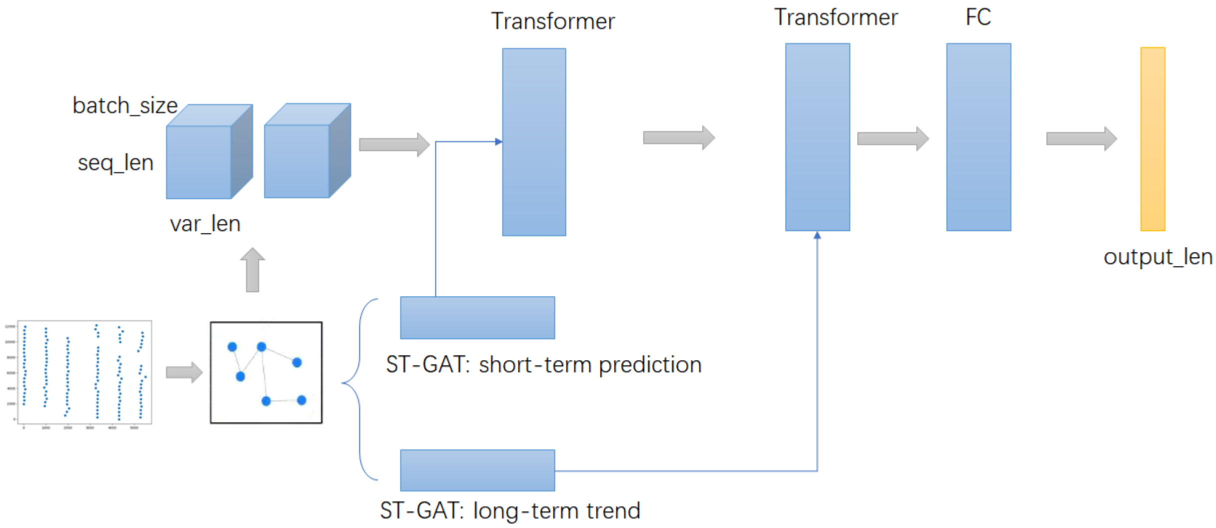

Figure 3: The structure of GCN-Transformer.

Table 2: Model performance in Baidu KDD Cup.

model	performance at final phase
GRU	46.96
Graph-former	47.15
GCN-Transformer	46.65
ensemble	46.34

6 CONCLUSION

We solve the problem of Wind Power Forecasting (WPF) by deep learning techniques. We propose a model based on graph convolutional networks with transformer. The proposed GCN-Transformer achieves a score of 46.65 at the final phase of Baidu KDD Cup 2022 under the framework of paddlepaddle. We also develop baseline models based on GRU unit. Finally, we achieve the score of 46.34 by the ensemble of three models.

REFERENCES

[1] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems* 29 (2016).

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[3] Linhao Dong, Shuang Xu, and Bo Xu. 2018. Speech-transformer: a no-recurrence sequence-to-sequence model for speech recognition. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 5884–5888.

[4] Marco Gori, Gabriele Monfardini, and Franco Scarselli. 2005. A new model for learning in graph domains. In *Proceedings. 2005 IEEE International Joint Conference on Neural Networks*, Vol. 2. 729–734.

[5] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. *Annual Conference on Neural Information Processing Systems* 30 (2017).

[6] Qinghua Hu, Shiguang Zhang, Man Yu, and Zongxia Xie. 2015. Short-term wind speed or power forecasting with heteroscedastic support vector regression. *IEEE Transactions on Sustainable Energy* 7, 1 (2015), 241–249.

[7] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*.

[8] Yann LeCun, Koray Kavukcuoglu, and Clément Farabet. 2010. Convolutional networks and applications in vision. In *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*. 253–256.

[9] Ron Levie, Federico Monti, Xavier Bresson, and Michael M Bronstein. 2018. Cayleynets: Graph convolutional neural networks with complex rational spectral filters. *IEEE Transactions on Signal Processing* 67, 1 (2018), 97–109.

[10] Shiyang Li, Xiaoyong Jin, Yao Xuan, Xiyou Zhou, Wenhu Chen, Yu-Xiang Wang, and Xifeng Yan. 2019. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *Advances in neural information processing systems* 32 (2019).

[11] Meng Liu, Zhengyang Wang, and Shuiwang Ji. 2021. Non-local graph neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).

[12] Shizhan Liu, Hang Yu, Cong Liao, Jianguo Li, Weiyao Lin, Alex X Liu, and Schahram Dustdar. 2021. Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In *International Conference on Learning Representations*.

[13] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. 2008. The graph neural network model. *IEEE Transactions on Neural Networks* 20, 1 (2008), 61–80.

[14] George Sideratos and Nikos D Hatziargyriou. 2007. An advanced statistical method for wind power forecasting. *IEEE Transactions on power systems* 22, 1 (2007), 258–265.

[15] Alessandro Sperduti and Antonina Starita. 1997. Supervised neural networks for the classification of structures. *IEEE Transactions on Neural Networks* 8, 3 (1997), 714–735.

[16] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[17] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. In *International Conference on Learning Representations*.

[18] Qingsong Wen, Tian Zhou, Chaoli Zhang, Weiqi Chen, Ziqing Ma, Junchi Yan, and Liang Sun. 2022. Transformers in time series: A survey. *arXiv preprint arXiv:2202.07125* (2022).

[19] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. 2021. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in Neural Information Processing Systems* 34 (2021), 22419–22430.

[20] Xu Yang, Hanwang Zhang, and Jianfei Cai. 2020. Auto-encoding and distilling scene graphs for image captioning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).

[21] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 11106–11115.

[22] Jingbo Zhou, Xinjiang Lu, Yixiong Xiao, Jiantao Su, Junfu Lyu, Yanjun Ma, and Dejing Dou. 2022. SDWPF: A Dataset for Spatial Dynamic Wind Power Forecasting Challenge at KDD Cup 2022. *Techincal Report* (2022).