A combination of Spatial-Temporal Graph Transformer Model and LSTM Model (Team: noritoshiTeam)

TAMURA,Noritoshi(Member:noritoshi)*
tamura@feg.co.jp
Financial Engineering Goup, Inc.
Tokyo, Japan

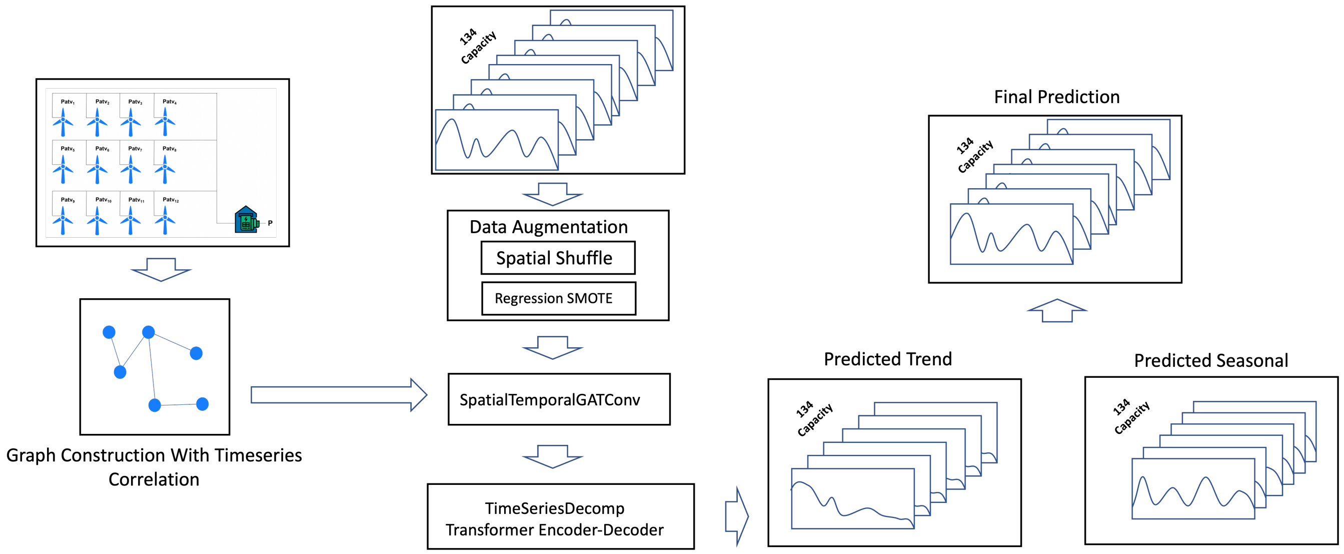

Figure 1: Spatial-Temporal Graph Transformer Model [4]

ABSTRACT

Wind Power Forecasting with Spatial-Temporal Graph Transformer Model [4] is a experimental approach method of spatial-temporal forecasting using Graph. It seems that Spatial-Temporal Graph Transformer Model has over-fitting tendency. By introducing taking the average of Spatial-Temporal Graph Transformer Model results and LSTM Model results,we are able to obtain improved accuracy of prediction.

ACM Reference Format:
TAMURA,Noritoshi(Member:noritoshi). 2022. A combination of Spatial-Temporal Graph Transformer Model and LSTM Model (Team: noritoshiTeam). In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (KDDCUP 2022).* ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

KDDCUP 2022, August 14–18, 2022, Washington DC
© 2022 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Wind Power Forecasting (WPF) aims to accurately estimate the wind power supply of a wind farm at different time scales. Wind power is a kind of clean and safe source of renewable energy, but cannot be produced consistently, leading to high variability. Such variability can present substantial challenges to incorporating wind power into a grid system. To maintain the balance between electricity generation and consumption, the fluctuation of wind power requires power substitution from other sources that might not be available at short notice (for example, usually it takes at least 6 hours to fire up a coal plant). Thus, WPF has been widely recognized as one of the most critical issues in wind power integration and operation. There has been an explosion of studies on wind power forecasting problems appearing in the data mining and machine learning community. Nevertheless, how to well handle the SD problem is still challenging, since high prediction accuracy is always demanded to ensure grid stability and security of supply[8]. Spatial-Temporal Graph Transformer Model is a new deep-learning approach which is able to solve complex problem such as treating a kind of Spatial-Temporal Data.

2 SOLUTION OVERVIEW

This solution[1] is consist of Spatial-Temporal Graph Transformer Model(STGT) [4] and simple LSTM[3] Model.Those components

[1] Codes in https://github.com/noritoshitamura/noritoshi_kddcup2022

Table 1: Statistics of the SDWPF data

Days	Interval	# of columns	# of turbines	# of records
245	10 minutes	13	134	4,727,520

are based on two types of official baseline models. Those Model have build and predict individually, and then as final prediction result ,take average of both predict values. SDWPF Dataset[8] is real data and relatively dirty than any other kind of general time series data, so considering Spatial-Temporal modeling approach is expecting valid as complementary effect to missing values.The LSTM Model has 134 of turbines model individually. Those are regarded as simple multivariate time series model. Using Graph to consider relation of neighboring turbines, it is possible that the model could accommodate accidental event records, such as overheat a turbine, down for maintenance or sensors broken etc.

3 DETAILED METHOD

The solution use Spatial-Temporal Graph Transformer Model(STGT) and LSTM Model.Two Models are build separately using different Features. At forecasting, simply calculate mean of those predictions values as final results.

$$Predictions = \frac{STGT\ predictions + LSTM\ predictions}{2} \tag{1}$$

3.1 SDWPF Dataset

The SDWPF dataset[8] is collected from the Supervisory Control and Data Acquisition (SCADA) system of a wind farm.SDWPF is obtained from the real-world data from Longyuan Power Group Corp. Ltd. (the largest wind power producer in China and Asia). The SCADA data 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 is shown in the Table 1 and its columns in table 2. For feeling the atmosphere of our mission, Fig.2 and Fig.3 are good samples of a snapshot of the SCADA data. that show Wind speed(Wspd) and Active Power(Patv) on turbines location at same time. Turbines location is specified with X and Y Coordinate. As a representative example, you can see that there are several groups of values, such as the difference between the left and right sides of the figures.

3.2 Spatial-Temporal Graph Transformer Model(STGT Model)

The Model is based on Autoformer[7] and incorporate Graph Learning Model[5] with PGL of PaddlePaddle Framework. It is natural to think that if many wind power turbines located in a area, turbines nearby activity have interaction or correlation each other. There is a possibility of the best approach of the SDWPF task solution. The model uses SMOTE regression with Graph and then decompose trend and seasonal factors by multi-head attention and re-compose from those variation at prediction. In Graph creataion, Nodes are 134 turbines and Edges is made from top five ranked Nodes that correlation of series of Patv values with, instead of considering each distance of location. Data augmentation is a basic technology in the area of image processing that often lack of observations for modeling. In this Model, as spatial shuffle, random number decide to permute Turbine's values or not($Probapility < 0.8$), and as regression SMOTE behavior of replacing values are weighted average of original turbine and other one that randomly picked, using weights determined random numbers that follow a beta distribution($\alpha = 0.5, \beta = 0.5$). SpatialTemporalGATConv that consists of spatial dimension of reduction depending on GAT(Graph Attension Networks) and temporal dimension of Convolusion depending on Conv1d, is utilized of Embedding layers of Transformer[6] components. TimeSeriesDecomp is component for the moving average to smooth out periodic fluctuations and highlight the long-term trends. For length-L input series $\chi \in \mathbb{R}^{L\times d}$ the process is:

$$\begin{aligned} \chi_t &= AvgPool(Padding(\chi)) \\ \chi_s &= \chi - \chi_t \end{aligned} \tag{2}$$

where $\chi_s, \chi_t \in \mathbb{R}^{L\times d}$ denote the seasonal and the extracted trend-cyclical part respectively.We adopt the AvgPool(·) for moving average with the padding operation to keep the series length unchanged.

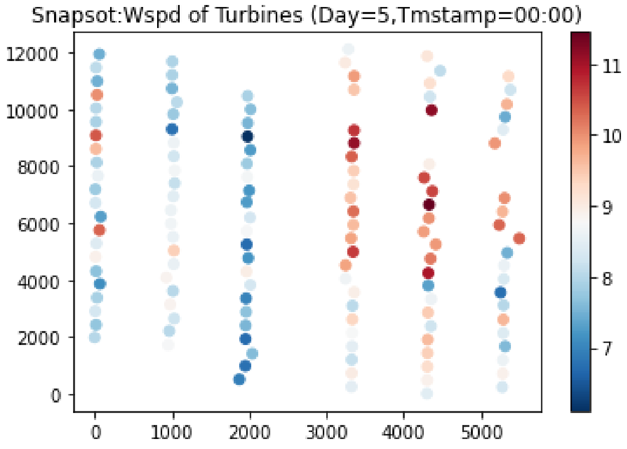

Figure 2: Example snapshot of Wspd by turbine locations.

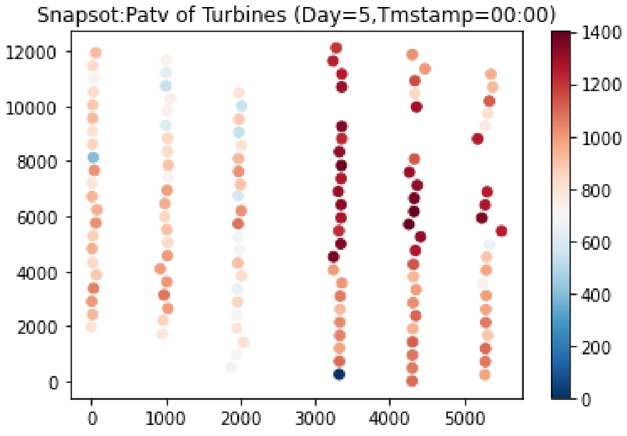

Figure 3: Example snapshot of Patv by turbine locations

Table 2: Column names and their specifications of the SDWPF data.

Column	Column Name	Specification
1	TurbID	Wind turbine ID
2	Day	Day of the record
3	Tmstamp	Created time of the record
4	Wspd (m/s)	The wind speed recorded by the anemometer
5	Wdir (°)	The angle between the wind direction and the position of turbine nacelle
6	Etmp (℃)	Temperature of the surounding environment
7	Itmp (℃)	Temperature inside the turbine nacelle
8	Ndir (°)	Nacelle direction, i.e., the yaw angle of the nacelle
9	Pab1 (°)	Pitch angle of blade 1
10	Pab2 (°)	Pitch angle of blade 2
11	Pab3 (°)	Pitch angle of blade 3
12	Prtv (kW)	Reactive power
13	Patv (kW)	Active power (target variable)

[7]. Transformer Encoder-Decoder components have two layers of Encoder, one layers of Decoder and 8 Heads. With these components and layers,combining forecasts decomposed into trends and seasonal variations.The model is almost same the official Spatial-Temporal Graph Transformer baseline Model[4] (Fig. 1). But the baseline code is made for just building a model, and dose not have enough codes for forecasting purpose, have to use prediction term data which has true values for getting time and weekday information. To forecast always from Test data, adding codes for making future terms of dummy data which has only times and weekday, increasing from last values of Test data. Using model features in the STGT Model are shown in Table 3.

Table 3: Features of STGT Model

Features	Description
Wspd	The wind speed
Ndir	Nacelle direction
Pab1~Pab3	Pitch angle of blades
Wdir	Wind relative direction
Wind_dir	wind absolute direction(*1)
Etmp	Environment Temp.
Tdiff	Difference between Inside and Env. Temp.(*2)
Prtv	Reactive power
Patv	Active Power

"Wind_dir" is absolute direction of wind (North is 0 degree) and calculation is,

(1) if Wdir is not missing calculate modular division by 360
(2) if Ndir is not missing calculate modular division by 360
(3) Take sum of both and then calculate modular division by 360 again

”Tdiff” is different Temp. between Inside and Environment Temp. for capturing generator unexpected warming and calculation is,

(1) Itmp and Etmp values change as limit between -10 and 80
(2) Above Itmp minus Etmp

Those Features with weekday(7 days cycle) and 10 min. intervals (144 steps cycle is a day) for auto-correlation depending on human activity (Considering Weekly and daily cycle) are used in the STGT Model. For reference,the official STGT baseline code use time intervals as 288 steps cycle, in other word,2 days in one cycle.

3.3 LSTM Model

The LSTM Model is change to LSTM(Long-Short Term Memory)[3] from GRU(Gated Recurrent Unit)[1] in the Simple GRU baseline Model[2]. It seems slightly better using LSTM than GRU for such complex time series data. The LSTM Model consist 2 LSTM blocks and trained with FilterMSEloss function Which used in the STGT Model for Considering a Generator has Unknown or abnormal Patv, with "onoff" flag of feature values. Using features in the LSTM Model are shown in table 4. As you know, ideal electric-generating capacity(P) of a turbine is able to calculate with following equation.

$$P = \frac{1}{2} A \rho v^3 \tag{3}$$

where A is effective area, ρ is air density and v is wind speed. Real electric-generating capacity has saturation. Therefore, when draw scatter plots of "Wspd" and "Patv" , we will see them distribute around S-shaped curve. In the features, "Area" is corresponding to A and "wp" is corresponding to v^3. ”Tdiff” is same name and concept of STGT Model’s one but limit value is different. Say,

(1) Itmp and Etmp values change as limit between -10 and 100
(2) Above Itmp minus Etmp

"onoff" is defined for unknown or abnormal values flag and used in FilterMSEloss function.

(1) Itmp and Etmp values change as limit between -10 and 100
(2) Above Itmp minus Etmp

3.4 Loss function

In Accordance with "Caveats about the data" section in the technical report of SDWPF[8], both models use FilterMSELoss function, to avoid Models learning from unknown and abnormal values. FilterMSEloss function normally returns MSE(Mean Squred Errors),

Table 4: Features of LSTM Model

Features	Description
Ndir	Nacelle direction
Pab1~Pab3	Pitch angle of blades
Area	= cosine(Wdir)
Itmp	Inside Temp.
Tdiff	Difference between Inside and Env. Temp.
wp	= Wspd **3
onoff	Flag of illegal or unknown condition
Prtv	Reactive power
Patv	Active Power

but if Patv would be recognized as Unknown or abnormal values on following conditions, the function returns 0.

Unknown values

- if at time t, $Patv \leq 0$ and $Wspd > 2.5$, then the actual active power Patv of this wind turbine at time t is unknown
- If at time t, $Pab1 > 89°$ or $Pab2 > 89°$ or $Pab3 > 89°$, then the actual active power Patv of this wind turbine at time t is unknown

Abnormal values

- if at time t, there are $Ndir > 720°$ or $Ndir < -720°$, then the actual active power Patv is wind turbine at time t is abnormal
- if at time t, there are $Wdir > 180°$ or $Wdir < -180°$, then actual active power Patv is wind turbine at time t is abnormal

4 EXPERIMENT

4.1 Hyper parameters and Scores

At training each Models, Model parameters were searched for the best by trial and error. Whole of the dataset have 245 days,used first 214 days as Train data, and remain 31 days for Vaildation in LSTM Model, or 16 days validation and 15 days test in the STGT Model. (Table 5.)

Table 5: How divide 245 days of Dataset

Model	Train	Valid	Test
STGT Model	214	31	-
LSTM Model	214	16	15

Data input length parameter of the STGT Model is 144, because it's not good that when increasing this parameter. It seems that predictor occasionally generates a few unique values. On the other hand, data input length parameter of the LSTM Model is 432, which is the best value in 144,288,432,1004 and 2008 that I tried. As a result, the short Length parameter of the STGT model tends to work on the spatial dimension, while the longer the LSTM model parameter is expected to work on the temporal dimension. Batch size is one of a key parameter that affects Model performance and accuracy. Batch size paramter is 24 on the STGT Model and is 45 on the LSTM Model.

After building both Models, make pair of predict values, then taking the averages of pairs.It is a way to get better score than individual predictions scoring result (Table 6). Final score is 43.9505 at local evaluation. Before getting final result, The average's score is slightly worse than the STGT Model's one, but I experieced invividual predcting with the STGT Model and the LSTM Model were same level of accuracy and using the average values of them was better Rank on phase 2 Leaderbord. I've get a good Rank early in Phase 3 Leaderboard (its local evaluation scores are shown in table 7). Using same Models, I add codes in preprocessing for limiting Wspd values upto 26.29, which is maximum value in provided period of SCADA data. We can see the improvement of the Average's score. It is noteworthy that the average score changed relatively significantly despite the small improvement in the STGT Model's score. In the other hand, it is interesting to note that if the limiting Wspd code added only to preprocessing of the STGT Model, the Average's score becomes -43.950916864, a little worse. I think that these phenomenon indicates that overfitting are likely to occur on not only STGT Model but also the LSTM Model, with SDWPF data. In Spatial-Temporal modeling as well, the problem of overfitting, which tends to occur in data with many missing values, can be avoided by ensemble models including Temporal type Models, leading to an improvement in final accuracy.

Table 6: Model Scores at local evaluation(My Best)

Model	Score
STGT Model	-43.784057996
LSTM Model	-45.096690815
Average of predict values	-43.950530428

Table 7: Model Scores at local evaluation(One before)

Model	Score
STGT Model	-43.784058096
LSTM Model	-45.093365072
Average of predict values	-43.953786771

4.2 Used Hardware,OS and Framework at Model building

The Models have build on PaddlePaddle Environment in a system show in table 8. With the system, learning the LSTM Model takes around 5 hours and a half, and learning the STGT Model takes around 2 hours.

Table 8: Used System at Model building

CPU	Ryzen7 3700X (Mother Board M450)
GPU	Nvidia Geforce GTX1080Ti
Memory	64 Gbyte
Storage	1 Tbyte M.2 SSD
OS	Ubuntu 20.04
Python	3.9.7
PaddlePaddle	2.2
PGL	2.1.5

REFERENCES

[1] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. https://doi.org/10.48550/ARXIV.1412.3555

[2] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2022. Smiple GRU Baseline Model. https://github.com/PaddlePaddle/PaddleSpatial/tree/main/apps/wpf_baseline_gru/. [Online; accessed 1-July-2022].

[3] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[4] Zhengjie Huang. 2022. Graph Neural Network by PGL. https://github.com/PaddlePaddle/PGL/tree/main/examples/kddcup2022/wpf_baseline/. [Online; accessed 16-May-2022].

[5] Junyoung Park. 2022. Wind Farm Power prediction with Graph Neural Network. https://aifrenz.github.io/present_file/wind_farm_presentation.pdf. [Online; accessed 16-May-2022].

[6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. https://doi.org/10.48550/ARXIV.1706.03762

[7] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. 2021. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. https://doi.org/10.48550/ARXIV.2106.13008

[8] 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).