AIStudio2338769Team: Long-Short Term Forecasting for Active Power of a Wind Farm

Wang Wenwei
wwangaw@connect.ust.hk
DAMO Academy, Alibaba Group
Hangzhou, Zhejiang, China

ABSTRACT

Accurate wind power forecasting is essential in coupling with its increasing penetration in power systems due to the high volatile nature. In this paper, we present our Long-Short Term Forecasting solution to the Spatial Dynamic Wind Power Forecasting Challenge at KDD Cup 2022. The task is to forecast 10 minutely wind power of 134 turbines from a wind farm for the next 48 hours, given the relative locations and internal status. We break the task to the nowcasting (0-3h) and short term (3h - 48h) part, targeting a more precise recent forecasting utilizing the inertia of wind and mean prediction in longer forecast horizons respectively. We use LightGBM only in the submission, along with several post processing tricks including a simple spatial ensemble. The source code is available at https://github.com/wenwei-pku/kddcup2022.

KEYWORDS

wind power forecasting, time series forecasting, gradient boost tree, Long Short-Term modeling, spatial-based ensemble

ACM Reference Format:

Wang Wenwei. 2018. AIStudio2338769Team: Long-Short Term Forecasting for Active Power of a Wind Farm. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Over the decades, wind power has played a substantial role in renewable energy sector. Due to the high fluctuations of wind at turbine height level, wind power usually reveals randomness, uncertainty intermittency and no clear pattern, as curated in Figure 1. The high penetration of wind power, thanks to the roaring installed capacity, has posed significant challenges on power systems. Therefore, accurate wind power forecasting is vital in system operation, dispatching and maintenance.

Numerous approaches on wind power forecasting has been proposed in recent years, among which the short term forecasting ranging from minute to a few days predominates. For example, day-ahead forecasting (24-48h) is essential in systems scheduling, while nowcasting (0-3h) helps in real-time grid operations and market clearing[1]. Considering the inertia of wind, nowcasting could be regarded as an extrapolation task of wind power time series hence short term time series forecasting techniques prevail at this stage. As forecast horizon grows, pure time series methods lose the track of wind pattern and Numeric Weather Prediction (NWP) takes over. NWP uses mathematical models of the atmosphere and oceans to predict the weather based on current weather conditions[1]. Derived from physical laws, NWP is able to provide relative accurate trend of wind and boost the accuracy of forecasting. It's widely recognized as one of the indispensable inputs in this scenario[2].

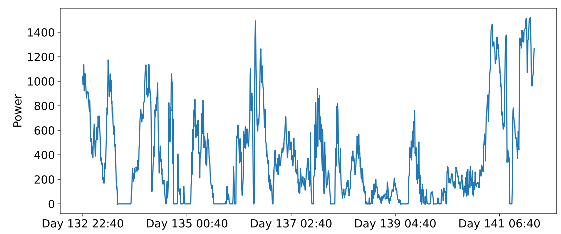

Figure 1: A slice of real power of Turbine 103. It shows the high volatile nature of wind power with no clear pattern.

KDD Cup 2022 is a combination of nowcasting and day-ahead forecasting. However, the absence of NWP makes it barely able to predict the real trend other than the mean. To address these challenges, we propose **Long-Short Term Forecasting** models, featuring a more precise prediction in the next a few hours and a plausible global average estimate in the long run.

The remainder of this paper is arranged as follows. Sections 2 presents the solution overview. Details are illustrated in section 3, followed by experimental results curated in section 4.

2 SOLUTION OVERVIEW

Our **Long-Short Term Forecasting** approach is based on two sources of information: **Mean** and **Inertia**. Since the competition's forecasting horizon is two days, we treat the first 3 hours (18 data points) and the remaining 45 hours (270 data points) separately as the inertia and the mean phase. We build two sets of gradient boost tree models (LightGBM [3]) for forecasts of each phase respectively, and LightGBM (LGB) is the only tool that we use. Though prolonging the inertia phase over 3 hours might further improve our result, the model size limits us from doing so. And we train an individual

[1] https://en.wikipedia.org/wiki/Numerical_weather_prediction

Conference'17, July 2017, Washington, DC, USA

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

model for each turbine to grab the slight difference between different turbines. Spatial information also greatly helps us through two simple and cheap **spatial-based ensemble** methods.

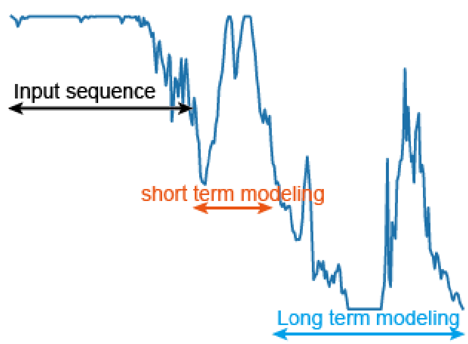

Figure 2: Long-Short Term Forecasting

3 DETAILED METHOD

The detailed method is presented in the following sections:

3.1 Data Preprocessing

The competition provides a dataset which retrieves from the SCADA system of a wind farm. It records each of the 134 turbines' operational data across 245 days. Besides the basic temporal information, 10 additional features (wind speed, wind direction, environmental and inside temperature, nacelle direction, pitch angle of 3 blades, reactiove power, active power) are also included.

The raw dataset has anomoly points, for instance, some extremely high values in Etmp and Wspd, which may disturb the training process. We preprocess the dataset in following steps:

- **remove outliers**. For Etmp and Wspd, we calculate the 25 and 75 percentile values among the whole dataset, replacing extra large or small values with *nan* and then using interpolation.
- **mask abnormal values**. We mark a sample as abnormal following rules in [4], i.e. if Wspd > 2.5 and Patv ≤ 0, or Pab $> 89°$ or Pab2 $> 89°$ or Pab3 $> 89°$, or Ndir $> 720°$ or Ndir $< -720°$, or Wdir $> 180°$ or Wdir $< -180°$. All the abnormal data are masked.

3.2 Feature Engineering

The overall feature engineering phase consists of raw feature preselection and selected feature engineering. There are 13 features in the raw training/inferencing dataset. TurbID, Day, and Tmstamp are kept as id and time indicator features. Based on the principles of physics and the unknown fluctuation nature of the prolonging features, we select only Etmp, Patv, and Wspd as base features for feature engineering. Following the relationship (1) between Wspd and Patv, we make another base feature Wspd^3.

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

We create three sets of features for the time series: Difference Features, Lagging features, and Rolling statistics.

Twenty-four Difference Features are created for Etmp, Patv, and Wspd each, which represent the feature difference between two consecutive data points across the 4 hours time horizon. Here we create 72 Difference features in total.

Lagging features are the 24 lagging values for Etmp, Patv, Wspd and Wspd^3 each, which stand for the raw feature in the closest 4 hours time horizon. Here we create 96 Lagging features in total.

Rolling statistics are calculated for Wspd, Wspd^3, Etmp and Patv, respectively. For the Rolling statistics, we calculate max, min, median, mean, std, skew as six statistics across rolling windows of time horizon 3,6,36,144 points. Here we create 96 Rolling statistics in total.

In addition, we create two hour-embedding features using sin and cos functions and a prediction indicator feature to indicate the prediction position among whole horizon 288.

In total we have 267 features for each sample.

Table 1: Features for Long short term modeling

Feature Type	Feature Count
Difference Features	72
Lagging Features	96
Rolling Statistics	96
Hour embedding feature	2
Prediction indicator	1

3.3 Long-term Wind Power Forecasting Model

The long-term forecasting model (LGB) has an output length of 288 points (2 days). Take $\mathbf{X}^k_{i-144\times14:i}$ as the input series of the past 14 days of k-th turbine, we call the feature engineering process to generate the features for the input of LGB model:

$$\mathbf{Z}^k_i = \mathrm{FeatureEngineer}(\mathbf{X}^{\mathrm{k}}_{\mathrm{i}-144\times14:\mathrm{i}}) \tag{2}$$

We concat ([:;:]) the feature $\mathbf{Z}^k_i$ and prediction indicator $j(j = 1, 2, ..., 288)$ to get the input of LGB model. y^k_j is the prediction of k-th turbine at j point.

$$y^k_j = \mathrm{LGB}([\mathbf{Z}^{\mathrm{k}}_{\mathrm{i}}; \mathrm{j}]) \tag{3}$$

The long-term model is expected to grasp the global pattern of the series. The wind power generated at wind turbines is strongly correlated to the wind speed (Wspd), however, the wind is a chaotic system, and the wind speed on a 2-day time scale is difficult to be modeled. When visualizing the output of the long-term model, we observe that the prediction is a quasi-horizontal line, which represents the global mean of the training data. We also test deep model (including TCN and LSTM) and similar output patterns are observed. Because of the limitation of model size, we employ 134 long-term models in total (one model for one turbine).

3.4 Short-term Wind Power Forecasting Model

The short-term forecasting model (LighGBM) is trained using the same feature engineering method with long-term model. The output length of the short-term model is 18. The short-term model consists of 18 sub-model:

$$y_j^k = \mathrm{LGB_j}(\mathrm{Z_i^k}) \tag{4}$$

where $j = 1, 2, ...18$. The j-th LGB sub-model is used for the prediction at j-th point (y_j^k). So we totally have 134×18 sub-models (one model for each turbine at each point). Based on the concept of inertial wind, the wind speed in short-term is predictable. We replace the first 18 points given by long-term models with that of short short-term models. This method will bring a consistent improvement when we have it tested offline.

3.5 Parameters Tuning

The hyperparameters of LGB models are determined with a 3-fold cross-validation on the whole dataset. We use customized hyperparameter for both short-term model and long-term model applied on each turbine. Finally, the models are trained with the whole 245-day dataset released.

The parameters for long-term LGB models are tuned for each turbine. We provided the training parameter for turbine 1 to 12 in the following Table 2, and the whole set of parameters for all turbines can be seen in the uploaded codes. The parameters for short-term LGB models are provided in Table 3.

Table 2: Long term model parameters

Parameter	value
learning_rate	0.03
boosting_type	gbdt
objective	regression_l2
metric	mae
num_leaves	63
min_data_in_leaf	100
feature_fraction	0.7
bagging_fraction	0.6
bagging_freq	5
min_split_gain	0.05
seed	16
n_estimators	80

3.6 Simple Spatial based Ensemble Tricks

Due to model size and running time constraints, we choose two simple, fast, and effective spatial based ensemble tricks.

The first one is the spatial closest turbine models ensemble: since we train individual model for each turbine and they are all built with the same features, we also use the most and the second closest turbines' models to forecast the wind power for each turbine. The final predicted wind power for each turbine at each time point are $Y(t) = 0.7 * M(X(t)) + 0.15 * M'(X(t)) + 0.15 * M''(X(t))$. The M,M',M" are trained model for each turbine and their closest and second closest turbine.

Table 3: Short term model parameters

Parameter	value
learning_rate	0.05
boosting_type	gbdt
objective	regression
metric	mae
num_leaves	40
min_data_in_leaf	60
feature_fraction	0.3
bagging_fraction	0.7
bagging_freq	5
max_depth	5
seed	16
num_iterations	64

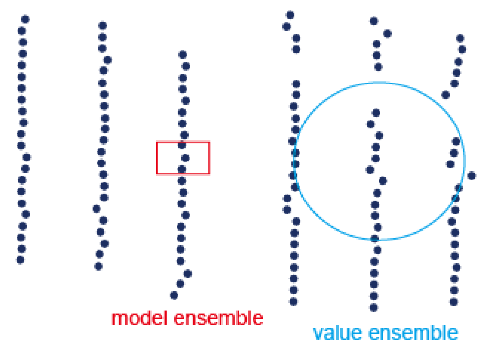

Figure 3: closest model ensemble and distance value ensemble

The second one is the spatial distance turbine value ensemble. This ensemble trick is based on the fact that most turbines produce similar wind power. Adding the mean of its neighbor turbine's predicted values will stabilize our prediction with almost no extra time costs. Here we select the neighbor turbines within 3k distance of our predicted turbine. The final predicted wind power for each turbine at each time point is $Y(t) = 0.5 * Y(t) + 0.5 * mean(Y_{neighbor}(t))$

3.7 Post Process

We applied a 1.08 multiplier factor for the 18 predicted points from the short-term forecasting model and a 1.18 multiplier factor for the resting 270 predicted points from the long-term model for each turbine. In Figure 4, we observe that the mean of Patv change according to the month (we give a fake 'month' to the training dataset). The model is trained on the 245-day training set and thus grasps the global mean of the whole dataset. However, the test set may have a different mean compared to the training set. So it is necessary to introduce a multiplier to fix the final prediction.

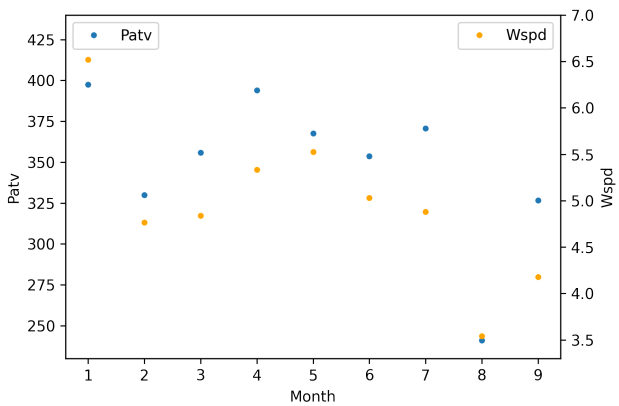

Figure 4: Mean of Patv and Wspd by month

4 EXPERIMENT PARTS

4.1 Offline Test Dataset

As the online submission quota is limited, we prepare an offline dataset to test the effectiveness of each model and trick.

The test dataset is generated according to the official report [4] about evaluation. We select 14 consecutive days and roll with a stride between 10 minutes and 900 minutes randomly, finally getting 21 instances.

4.2 Experiment results

The average score of 21 instances is recorded. We compare the result of baseline and with several optimization and tricks in Table 4.

Table 4: Model performance under different settings on summer test data. Each row builds on all the previous rows.

Model setting	offline test score	improvement
fixed parameters	-49.07	-
+ short term model	-48.11	0.96
+ tuned parameters	-48.05	0.06
+ spatial based ensemble	-47.95	0.10
+ whole ratio*1.18	-47.02	0.93
+ first18 ratio*1.08	-46.95	0.07

The baseline model is a global LGBRegressor with fixed parameters reported in Table 3 and it results a local score of 49.07. A global model tends to capture the mean value of the long forecasting period, and is less appropriate to short term forecasts. Introducing the set of Short-term forecasting models significantly improves the score by 0.96. Parameter tuning further further enhances our performance by 0.96.

As nearby turbines usually generate similar amounts of patvs, we propose the spatial based ensemble model and it modifies the result by 0.1. Observing that the mean patv fluctuates in seasons, we adjust the magnitude of our forecasts, which totally helps to raise the score by 1.

Figure 5 demonstrates predictions under various boosting models and tricks. The blue line is the ground truth, and the first 144 points is taken as the input history data. The orange line is the predictions with fixed parameters, which shows almost a mean value. The green line is the predictions with short models and shows a better performance on the first 10 time steps. The spatial based ensemble strategy gives the more stable red line prediction, and the post process with a multiplier factor makes the prediction closer to the true mean value.

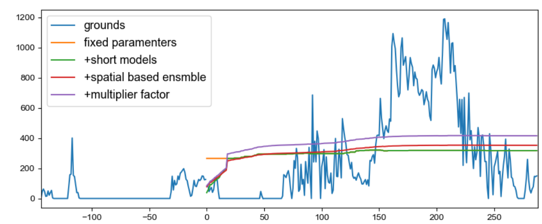

Figure 5: Predictions under boosting models and tricks

Our offline experiments indicates the effectiveness of the methods and tricks introduced in Section 3. We test some combinations online and receive the results in Table 5.

Table 5: Online test socres. a: fixed parameters, b: tuned parameters, c: short term models, d: spatial based ensemble, e: whole ratio*1.18, f: first18 ratio*1.08.

Model setting	online test score
a+b	-45.79
a+b+e	-45.50
b+c+e	-45.42
b+c+d+e	-45.29
b+c+d+e+f	-45.27

5 CONCLUSION

In this paper, we adopt a Gradient Boosting Tree based model for the wind power forecasting for next 48 hours. We build a short-term forecasting model to capture the inertia of wind and a long-term model to grasp the global mean. Our framework (LGB as backbone model combined with several post process tricks) achieved the 9th prize in Kddcup 2022.

REFERENCES

[1] M.Uzunoglu A.Tascikaraoglu. A review of combined approaches for prediction of short-term wind speed and power. *Renewable and Sustainable Energy Reviews*, 34:243–254, 2014.
[2] X.; Lin Z.; Lotfian S Hanifi, S.; Liu. A critical review of wind power forecasting methods—past, present and future. *Energies*, 13:3764, 2020.
[3] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30:3146–3154, 2017.
[4] Jingbo Zhou, Xinjiang Lu, Yixiong Xiao, Jiantao Su, Junfu Lyu, Yanjun Ma, and Dejing Dou. Sdwpf: A dataset for spatial dynamic wind power forecasting challenge at kdd cup 2022. *Techincal Report*, 2022.