KDD CUP
2022

WPFormer: A Spatio-Temporal Graph Transformer with Auto-Correlation for Wind Power Prediction

Xuefeng Liang[1], Qingshui Gu[1], Su Qiao[2], Zhuwang Lv[2], Xin Song[3]
[1] Tsinghua University, [2] Zhejiang Gongshang University, [3] University of Toronto

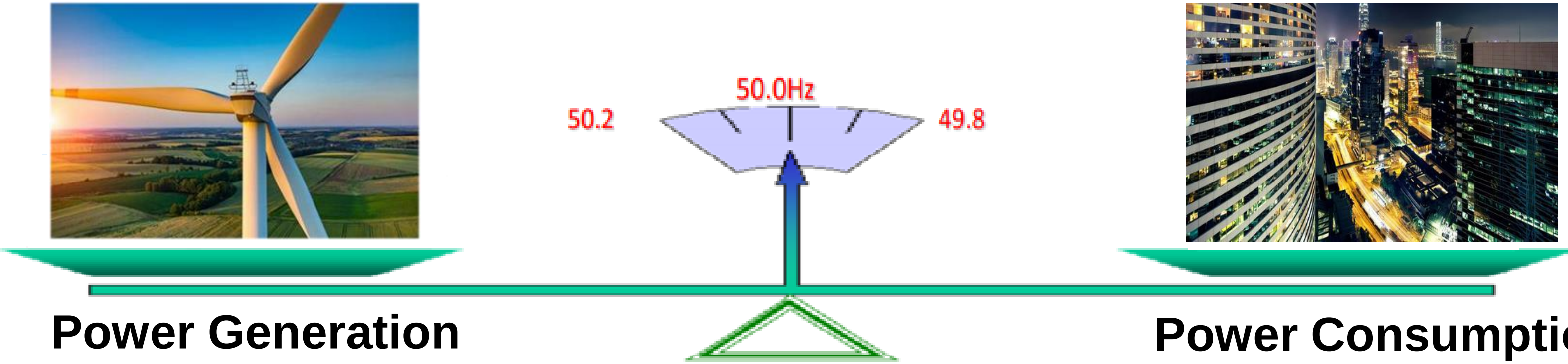
August 09, 2022

Team Zhanshijin

目录

CONTENTS

1. BACKGROUND



Power Generation

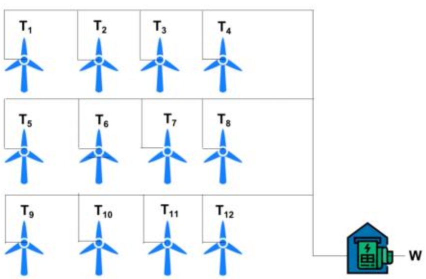
Power Consumption

Spatial distribution

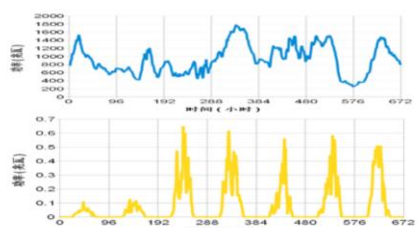
SDWPF

Abnormal data

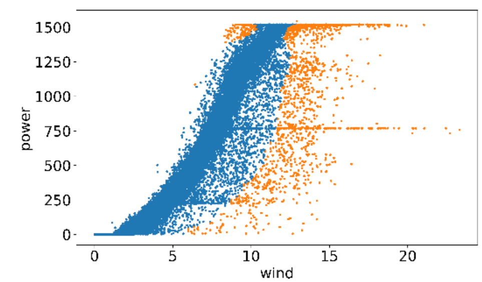
Dynamic context



The relative locations of all wind turbines given a wind farm for modeling the spatial correlation.



The weather situations and internal turbine status detected by each wind.



Since the data comes from real data, there are a large number of outliers.

✓ **The volatility of power generation and the long forecast time are huge challenges for this mission.**

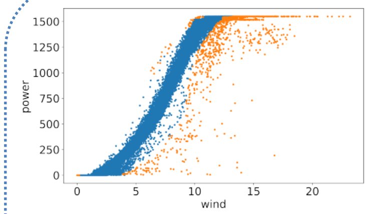
2. METHODOLOGY

- 2.1 Data Engineering Based on Wind Power Curves
- 2.2 Feature Engineering
- 2.3 WPFormer Structure
- 2.4 Point-by-point prediction based on tree model (POPtree)
- 2.5 Model Fusion Strategy
- 2.6 Exploratory Methods

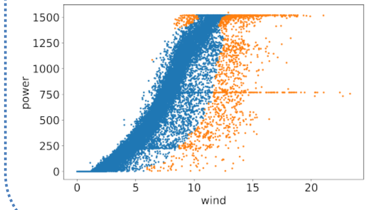
2.1 Data Engineering Based on Wind Power Curves

✓ Data Analysis

Data Type	Proportion(%)	Contents
Missing values	1.05	NULL
Zero values	26.72	$Patv < 0$ or $Prtv < 0$
Unknown values 1	6.33	$Patv \leq 0$ and $Wspd > 2.5$
Unknown values 2	20.83	$Pab1 > 89^\circ$ or $Pab2 > 89^\circ$ or $Pab3 > 89^\circ$
Abnormal values	22.89	$Ndir > 720^\circ$ or $Ndir < -720^\circ$ or $Wdir > 180^\circ$ or $Wdir < -180^\circ$
Mask values	22.89	The union of missing values, unknown values and Abnormal values.
Total values	28.64	The union of all the above.



(a) Turbine No. 1.



(b) Turbine No. 2.

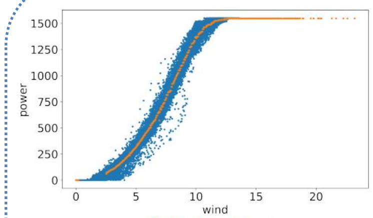
✓ Anomaly Detection

Use Isolation Forest algorithm to detection anomaly data.

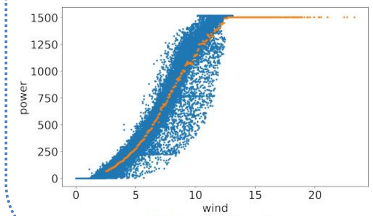
$$\boxed{P} = \frac{1}{2} C_p \rho A \boxed{v^3}$$

P—power(W) ρ—air density(kg/m^3) Cp—wind energy utilization coefficient

v—wind speed(m/s) A—impeller swept area(m^2)



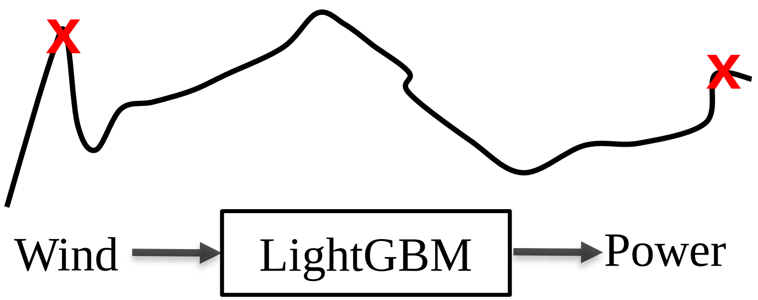
(a) Turbine No. 1.



(b) Turbine No. 2.

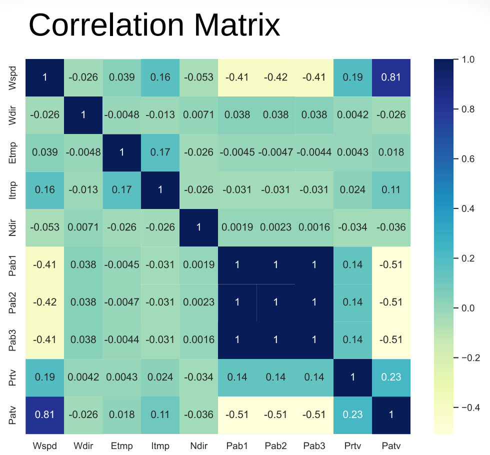
✓ Data Cleaning

Use LightGBM to repair anomaly data.



2.2 Feature Engineering

✓ Feature Selection



+

LightGBM Feature Selection

Selected Features:

Wspd, Pab1, Pab2, Pab3, Prtv

Excluded Features:

Tmstamp, Wdir, Ndir, Itmp

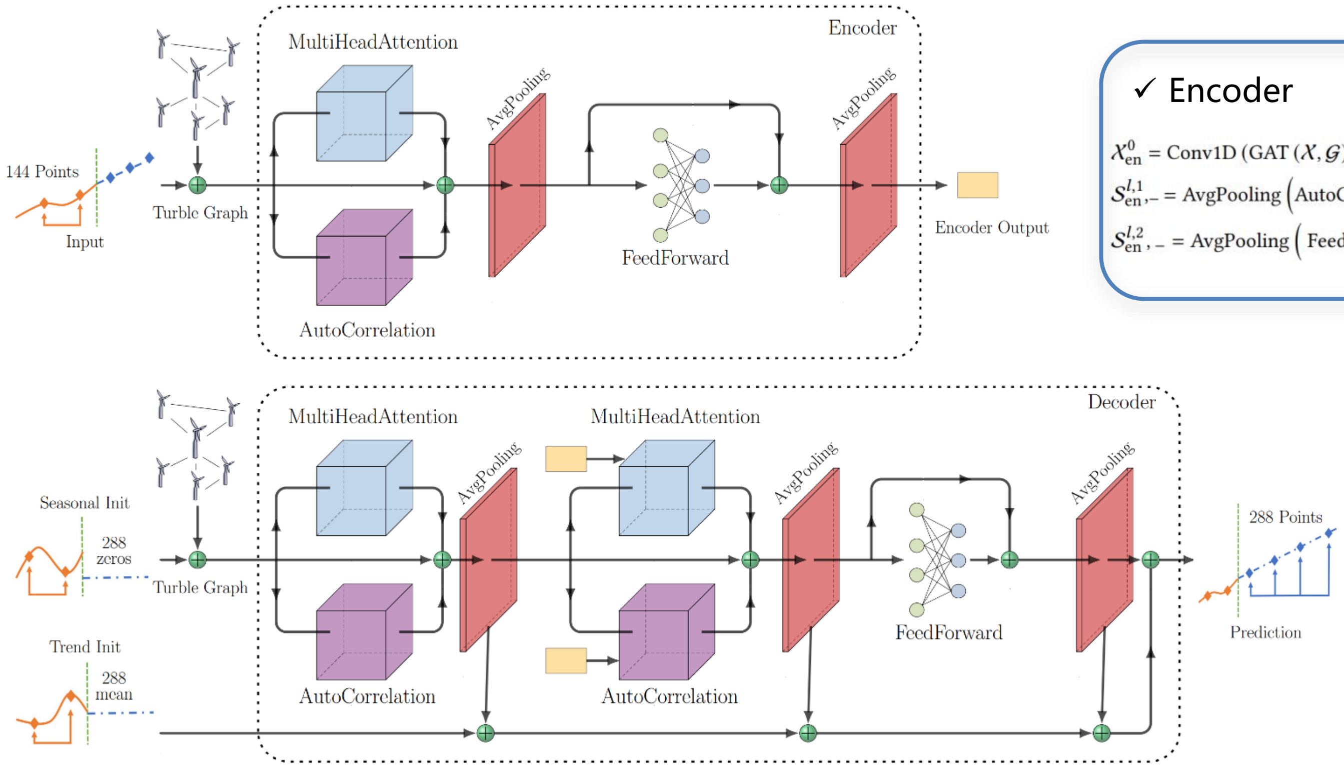
✓ Statistical Features

No.	Implication	No.	Implication
1	mean	2	standard deviation
3	maximum	4	skewness
5	minimum	6	kurtosis
7	absolute energy	8	autoregressive coefficient
9	autocorrelation	10	approximate entropy
11	c3 statistics	12	complex of time series

No.	Implication
13	quantile (0.1, 0.25, 0.5, 0.75, 0.9)
14	the sum over the absolute value of consecutive changes.
15	augmented Dickey-Fuller test
16	the number of values in the time series that are higher than the mean of the time series.
17	the length of the longest consecutive subsequence in the series that is bigger than the mean of the time series.
18	the length of the longest consecutive subsequence in the series that is smaller than the mean of the time series.
19	the mean over the absolute differences between subsequent series values.
20	the number of peaks in the time series.
21	the sum of all data points that are present more than once.

Statistical features (7)-(12) (14)-(21) refer to tsfresh.

2.3 WPFormer Structure



✓ Encoder

$$\mathcal{X}_{en}^{0} = \mathrm{Conv1D}\left(\mathrm{GAT}\left(\mathcal{X}, \mathcal{G}\right)\right),$$

$$\mathcal{S}_{en,-}^{l,1} = \mathrm{AvgPooling}\left(\mathrm{AutoCorrelation}\left(\mathcal{X}_{en}^{l-1}\right) + \mathcal{X}_{en}^{l-1}\right),$$

$$\mathcal{S}_{en,-}^{l,2} = \mathrm{AvgPooling}\left(\mathrm{FeedForward}\left(\mathcal{S}_{en}^{l,1}\right) + \mathcal{S}_{en}^{l,1}\right),$$

✓ Decoder

$$\mathcal{X}_{de}^{0} = \mathrm{Conv1D}(\mathrm{GAT}\left(\mathrm{ConCat}\left(\mathrm{AvgPooling}\left(\mathcal{X}\right), \mathcal{S}_{init}\right), \mathcal{G}\right)),$$

$$\mathcal{T}_{de}^{0} = \mathrm{ConCat}\left(\mathrm{AvgPooling}\left(\mathcal{X}\right), \mathcal{T}_{init}\right),$$

$$\mathcal{S}_{de}^{l,1}, \mathcal{T}_{de}^{l,1} = \mathrm{AvgPooling}\left(\mathrm{AutoCorrelation}\left(\mathcal{X}_{de}^{l-1}\right) + \mathcal{X}_{de}^{l-1}\right),$$

$$\mathcal{S}_{de}^{l,2}, \mathcal{T}_{de}^{l,2} = \mathrm{AvgPooling}\left(\mathrm{AutoCorrelation}\left(\mathcal{S}_{de}^{l,1}, \mathcal{X}_{en}^{N}\right) + \mathcal{S}_{de}^{l,1}\right),$$

$$\mathcal{S}_{de}^{l,3}, \mathcal{T}_{de}^{l,3} = \mathrm{AvgPooling}\left(\mathrm{FeedForward}\left(\mathcal{S}_{de}^{l,2}\right) + \mathcal{S}_{de}^{l,2}\right),$$

$$\mathcal{T}_{de}^{l} = \mathcal{T}_{de}^{l-1} + \mathcal{W}_{l,1} \cdot \mathcal{T}_{de}^{l,1} + \mathcal{W}_{l,2} \cdot \mathcal{T}_{de}^{l,2} + \mathcal{W}_{l,3} \cdot \mathcal{T}_{de}^{l,3}$$

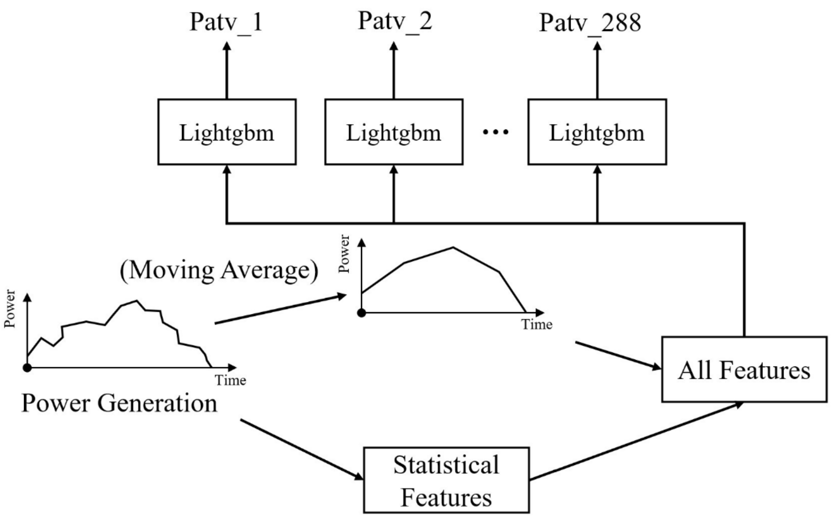
✓ Auto-Correlated Mechanism

It mainly calculates the sequence autocorrelation coefficient, aggregates similar subsequence information, realizes sequence-level connection, and completes better information aggregation.

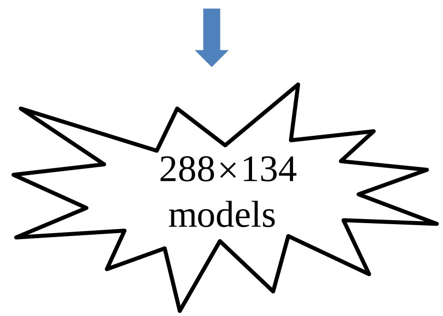
✓ Graph Neural Network Representation

Each turbine is a graph node. For a particular node, calculate the K nodes with the highest correlation with it, and establish connection edges.

2.4 POPtree

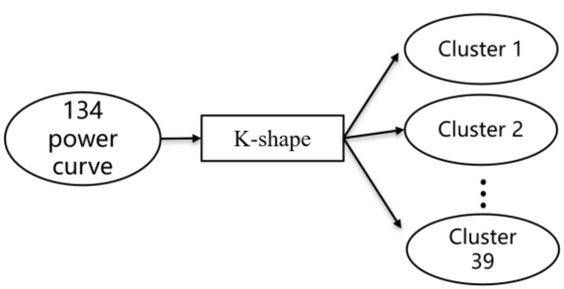


Each time point on each turbine power curve is predicted by one LightGBM model



✓ Reduce number of models (388 × 134 => 39 × 72)

a) Time series cluster and use same model in one cluster.



b) Down-sampled power generation at time scale

✓ Auxiliary Turbines

a) Calculate the Pearson coefficient between the power by each turbine.

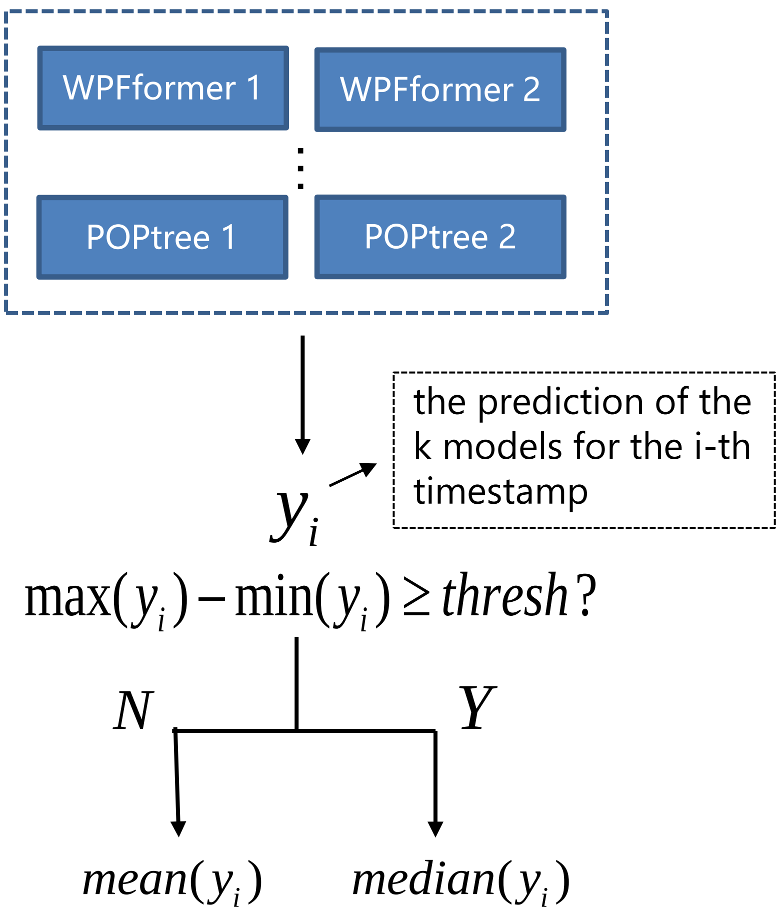
$$R_{ij} = \frac{cov(P_i, P_j)}{\sqrt{var(P_i) * var(P_j)}}$$

b) The top k turbines that are most similar to a turbine are defined as auxiliary turbines.

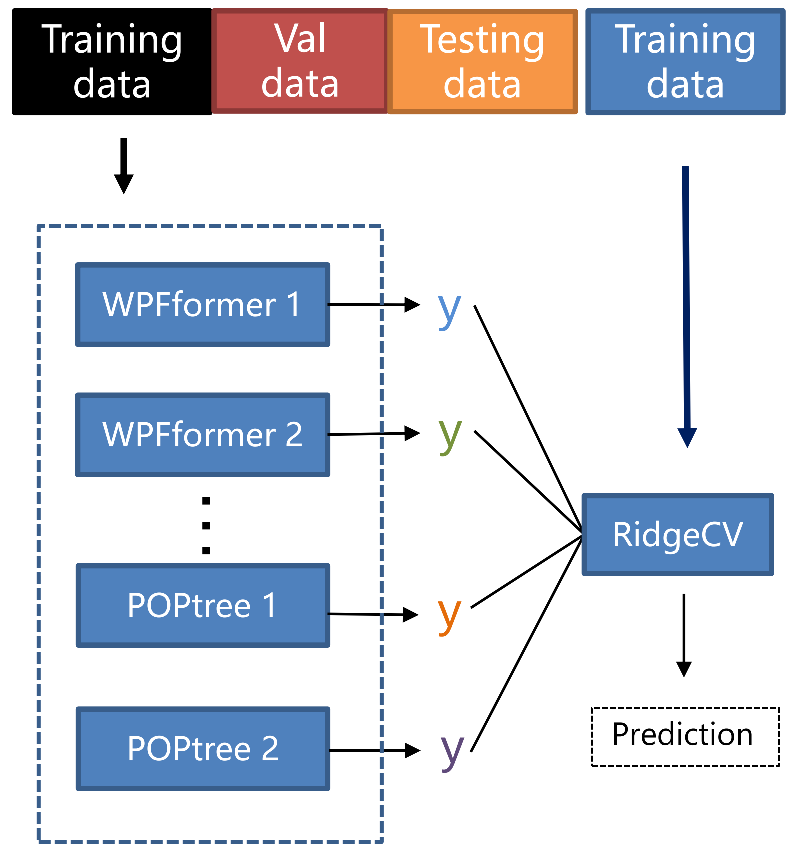
c) Average the power by a turbine and its auxiliary turbines.

2.5 Model Fusion Strategy

✓ **Hybrid Fusion**



✓ **Weighted Fusion**



The first method is to use a hybrid fusion method combining both mean and median.

Another is to use the linear weighting method to ensemble the models.

2.6 Exploratory Methods

✓ **Mixed Loss**

$$L_{mse} = \frac{1}{2k}\sum_{i=0}^{k}(y_i^{'} - y_i)^2$$

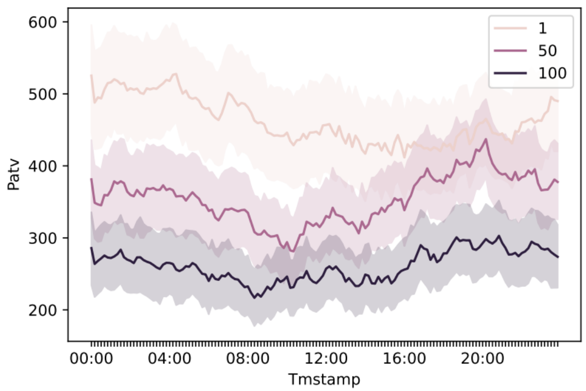
$$+$$

$$L_{l1} = \frac{1}{k}\sum_{i=0}^{k}\left|y_i^{'} - y_i\right|$$

$$L_{total} = L_{mse} + L_{l1}$$

Since the competition metric is the average of rmse and mae, so we also try to use other losses, such as L1loss, mseloss.

✓ **Models Based on Mean Learning**



$$P_{i,k} = \alpha_k \hat{E}(P_i) + \beta_k$$

Specifically, the model takes the average power and monitoring data of the previous 14 days of a particular turbine as input and then predicts the offset and scale for the next two days.

3. EXPERIMENTS

✓ Results of Phase 2

Serial No.	Model Components	Online Score	Offline Score
1	baseline	-	-46.83
2	phase 2 basic	-44.72740	-43.78076
3	+ feature screening1	-44.60840	-43.59530
4	+ model fusion	-44.58287	-43.63643
5	+ feature screening2 + data cleaning	-44.46343	-43.36928
6	+ Change the training parameters	-44.40641	- 43.26067
7	+ feature screening3	-44.39223	-43.28473
8	+ model fusion + Change the training parameters	-44.37068	-43.31144
9	+ POPtree	**-44.32841**	-43.28473

✓ Results of Phase 3

Serial No.	Model Components	Online Score	Offline Score
1	phase 3 basic	-45.48641	-43.33292
2	+ data cleaning	-45.28938	-43.31111
3	+ feature screening	-45.25772	-43.40196
4	+ autocorrelation mechanism	-45.20941	-43.10517
5	+ model fusion	**-45.13867**	-43.10917
6	Add the POPtree based on model No.2	-45.22724	43.35802

We have achieved **steady improvements** online using data cleaning, feature engineering, different models, training tricks, and fusion strategies.

4. CONCLUSIONS

- We first design data cleaning and correction based on the wind power curve and complete feature screening.
- Then we develop two models specifically for wind power scenarios, WPFormer and POPtree, respectively.
- Furthermore, we propose two effective model fusion methods to ensemble these two models.
- Eventually, the steady improvement of online scores shows the excellent performance of our solution.

5. REFERENCES

[1] Amirhossein Ahmadi, Mojtaba Nabipour, Behnam Mohammadi-Ivatloo, Ali Moradi Amani, Seungmin Rho, and Md. Jalil Piran. 2020. Long-Term Wind Power Forecasting Using Tree-Based Learning Algorithms. IEEE Access 8 (2020), 151511–151522. https://doi.org/10.1109/ACCESS.2020.3017442

[2] 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.

[3] Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. 2022. FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting. In Proceedings of the 39th International Conference on Machine Learning, Vol. 162. 27268–27286.

[4] 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.

[5] Shereen Elsayed, Daniela Thyssens, Ahmed Rashed, Lars Schmidt-Thieme, and Samer Hadi Jomaa. 2021. Do We Really Need Deep Learning Models for Time Series Forecasting? (2021).

[6] 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).

[7] Umut Firat, Seref Naci Engin, Murat Saraclar, and Aysin Baytan Ertuzun. 2010. Wind Speed Forecasting Based on Second Order Blind Identification and Autoregressive Model. In 2010 Ninth International Conference on Machine Learning and Applications. 686–691. https://doi.org/10.1109/ICMLA.2010.106

[8] Qinkai Han, Fanman Meng, Tao Hu, and Fulei Chu. 2017. Non-parametric hybrid models for wind speed forecasting. Energy Conversion and Management 148 (2017), 554–568.

[9] Xiaohui He, Ying Nie, Hengliang Guo, and Jianzhou Wang. 2020. Research on a Novel Combination System on the Basis of Deep Learning and Swarm Intelligence Optimization Algorithm for Wind Speed Forecasting. IEEE Access 8 (2020), 51482–51499.

[10] Yun Ju, Guangyu Sun, Quanhe Chen, Min Zhang, Huixian Zhu, and Mujeeb Ur Rehman. 2019. A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting. Ieee Access 7 (2019), 28309–28318.

THANKS FOR YOUR ATTENTION!