Baidu KDD CUP 2022

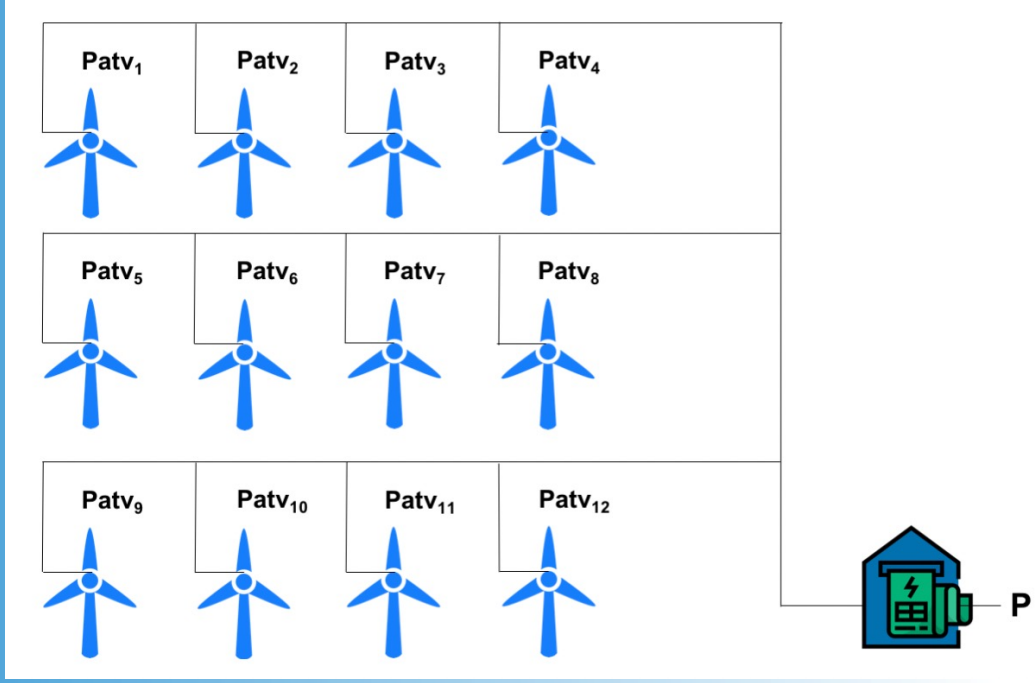
Spatial Dynamic Wind Power Forecasting

trymore： Regular track winner #2 & PaddlePaddle Track winner #1

Reporter Hanhan Liu
Sieyuan Electric Co., Ltd

CONTENTS

Introduction - Task

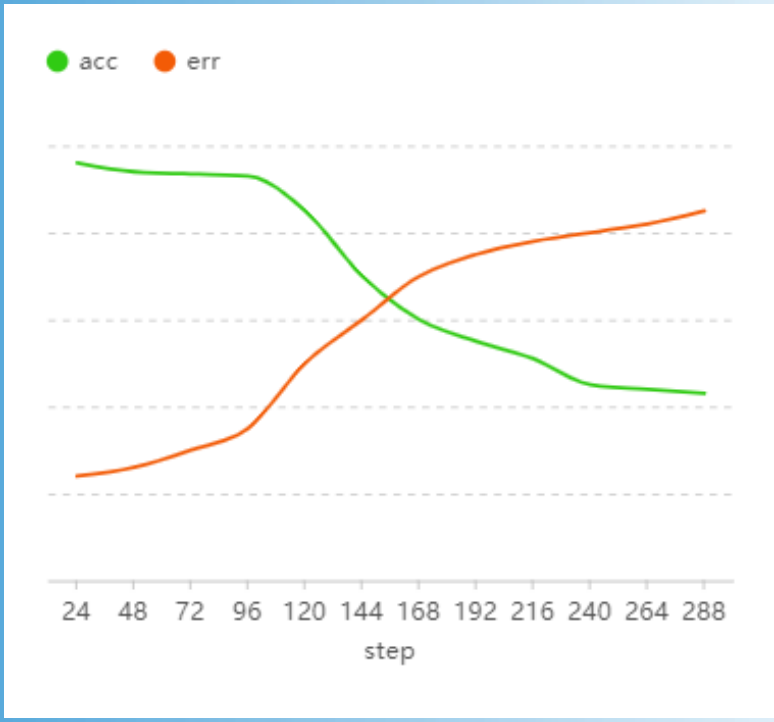


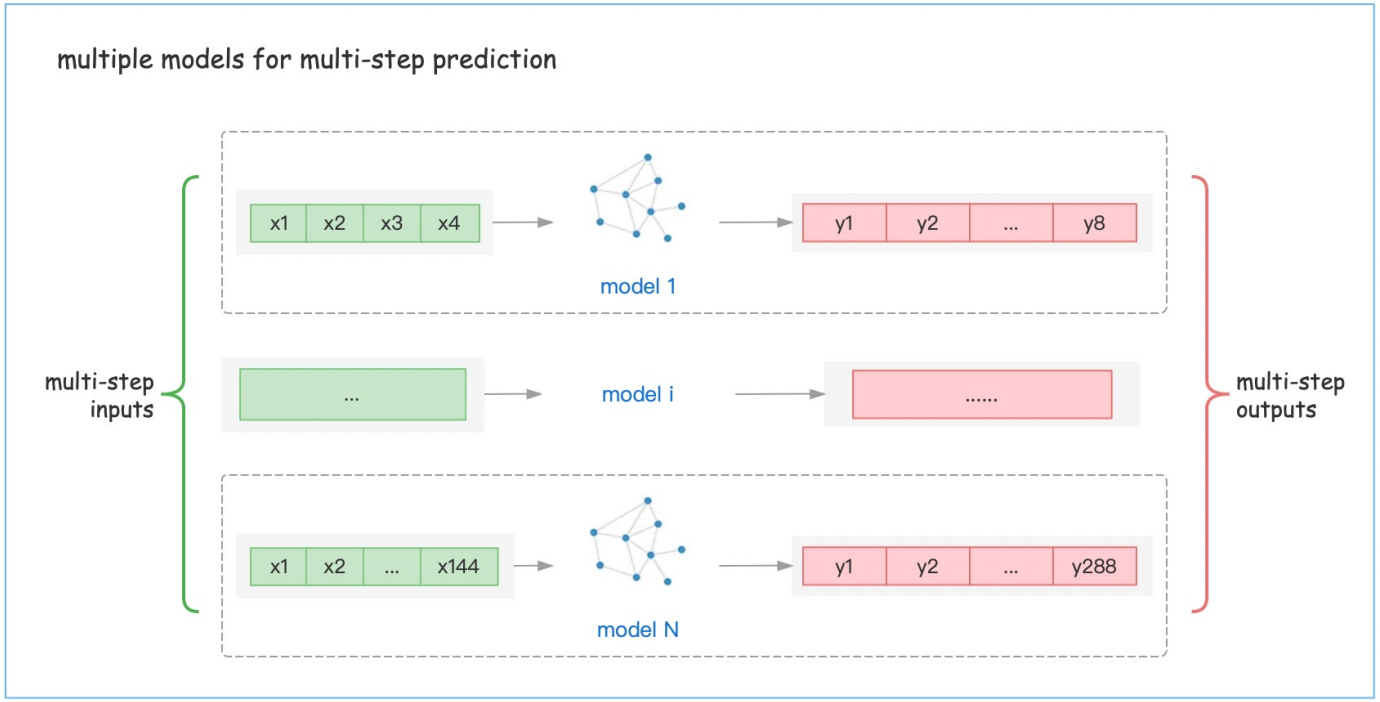
- The Baidu KDD Cup 2022 task is to accurately address the Spatial Dynamic Wind Power Forecasting ahead of 48 hours (It is required to output the predict values every 10 minutes).
- **Dataset:** Spatial distribution & Dynamic context.
- **Metric:** the average of RMSE and MAE.
- **Challenges:** outliers data, key features, accuracy prediction over time

Sieyuan

Solution - Multi-Step Prediction

- The longer the wind power output prediction time, the lower the accuracy will be.
- We use multiple models to make multi-step forecasts, rather than one-off forecasts.

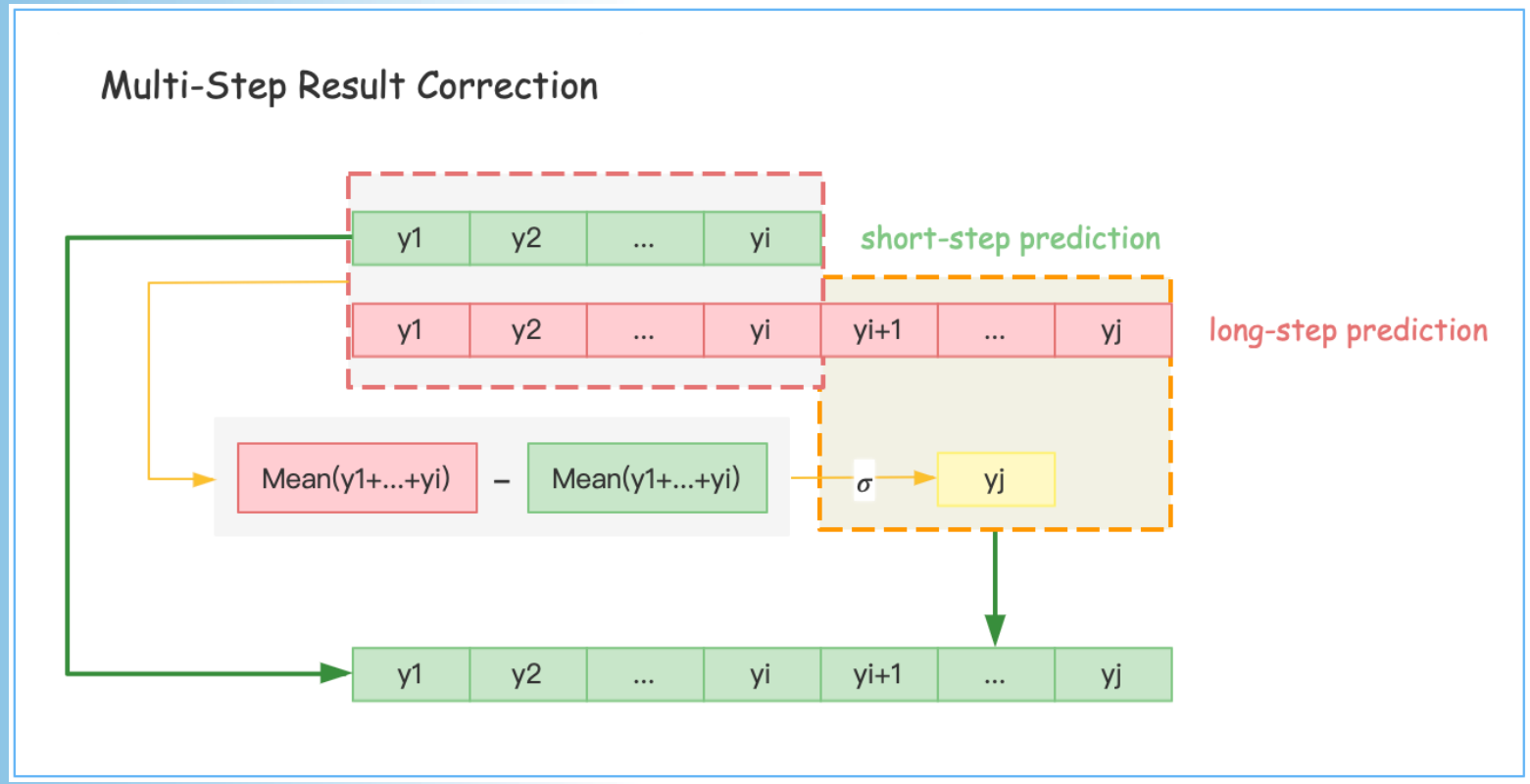




Sieyuan

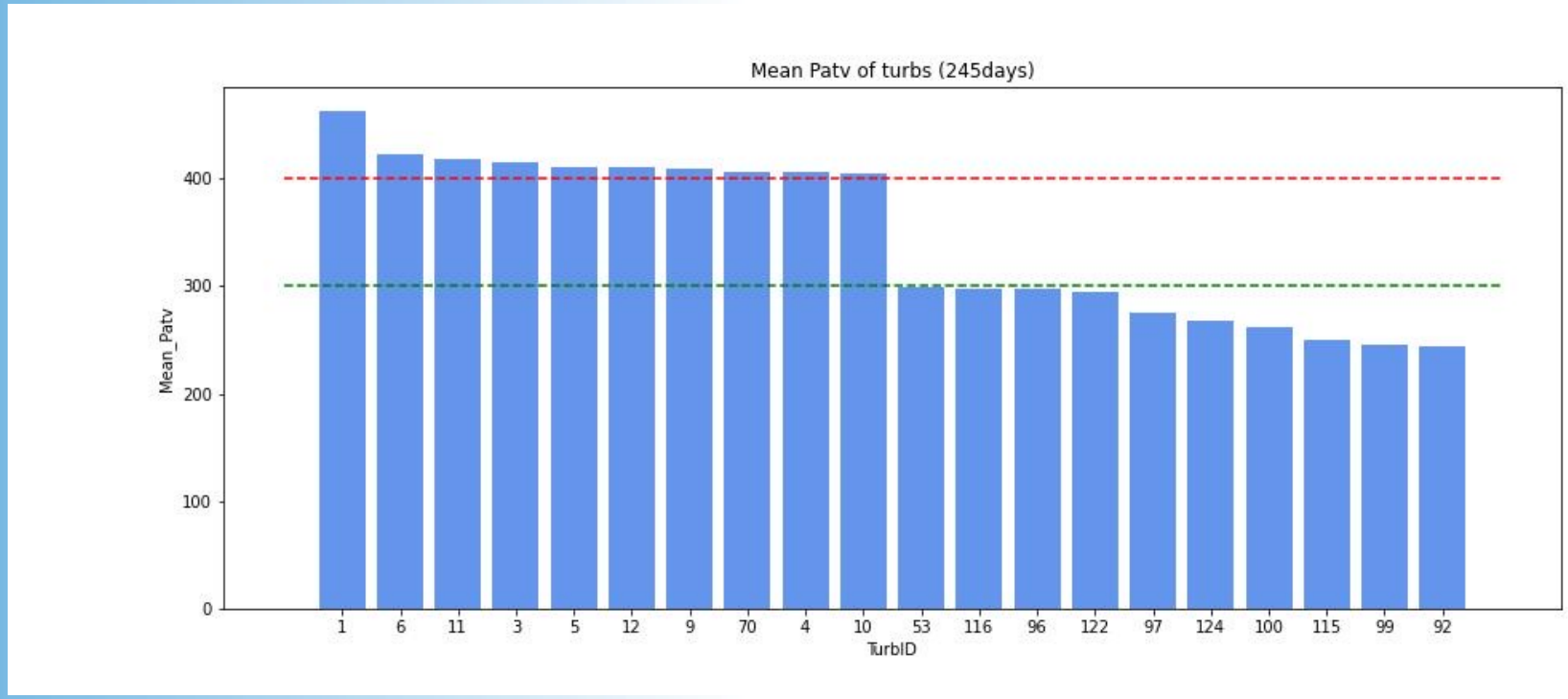
Solution - Multi-Step Result Correction

- The prediction output based on the shorter time range is more accurate.
- So, we use the short-range prediction results to correct the long-range prediction results.



Solution - Top N Turbines of Patv

- The higher average wind power (Patv) the turbine output, indicates that the output is closer to its theoretical output power.
- So we select top N wind turbines of highest Patv, and use their history data to train models to predict. After that, we will use the average of the inference result as the results of other turbines in the whole wind farm.



Sieyuan

Solution - Multiple Feature Engineering

- different feature engineering applied by different models

Multi-Step Tree Module

0. abnormal data filtering
1. add new feature:'hour_sin', 'hour_cos', 'cross'
2. feature lag (n=6)
3. feature diff (n=6)

Top-N Module

1. correction of outliers: 'Nidr', 'Wdir', 'Etmp', 'Itmp', 'Pab1', 'Pab2', 'Pab3'
2. normalization

Multivariant Fusion Module

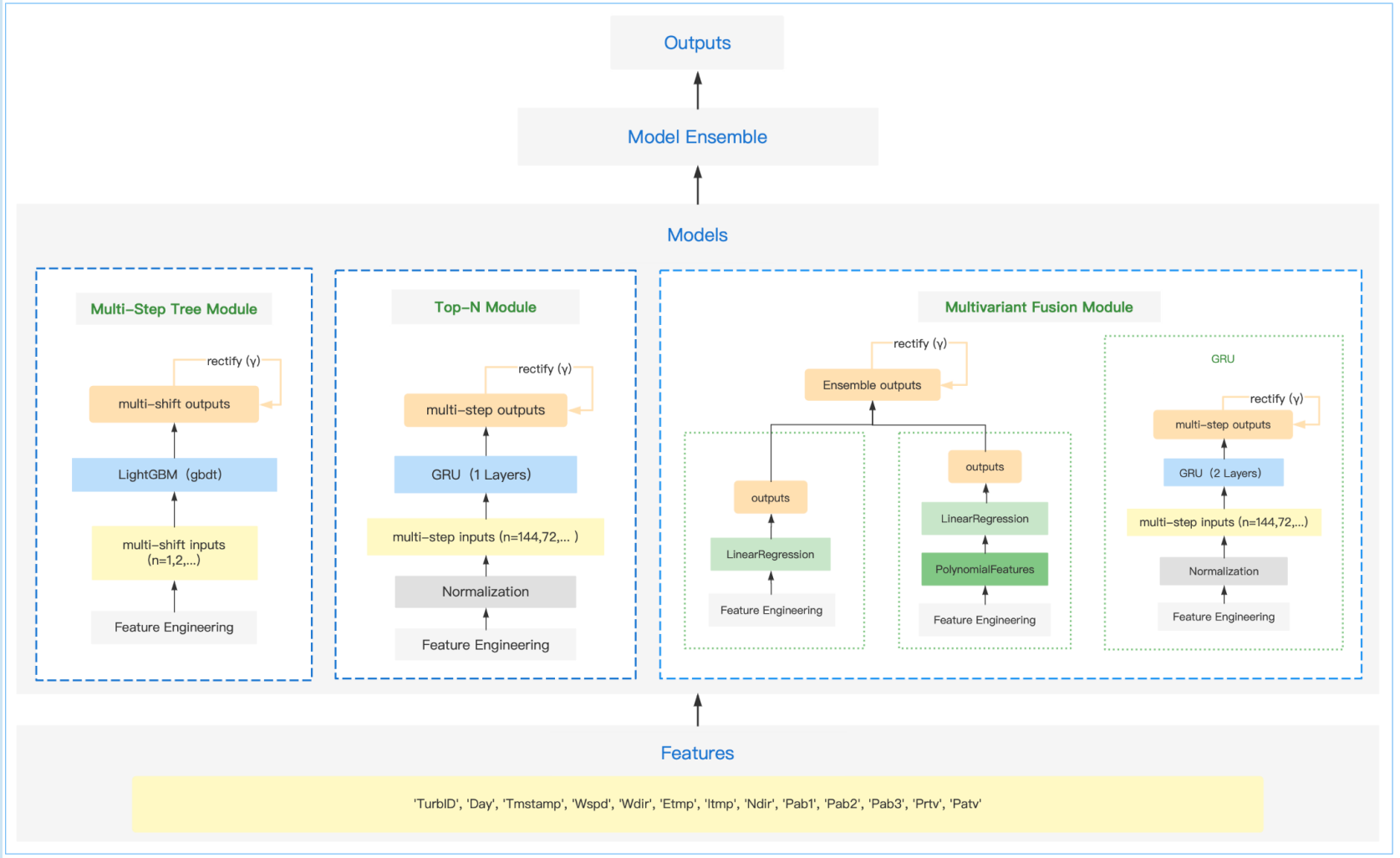
1. add new feature: 'Etmp_Max', 'TurbID__Etmp_Max__1008__max'
2. TurbID embeding: one-hot
3. normalization (standardize)

0. abnormal data filtering
1. add new feature: 'hour_sin', 'hour_cos', 'cross', 'Etmp_Max', ...
2. feature lag (n=12)
3. feature diff (n=12)

Sieyuan

Method - Framework

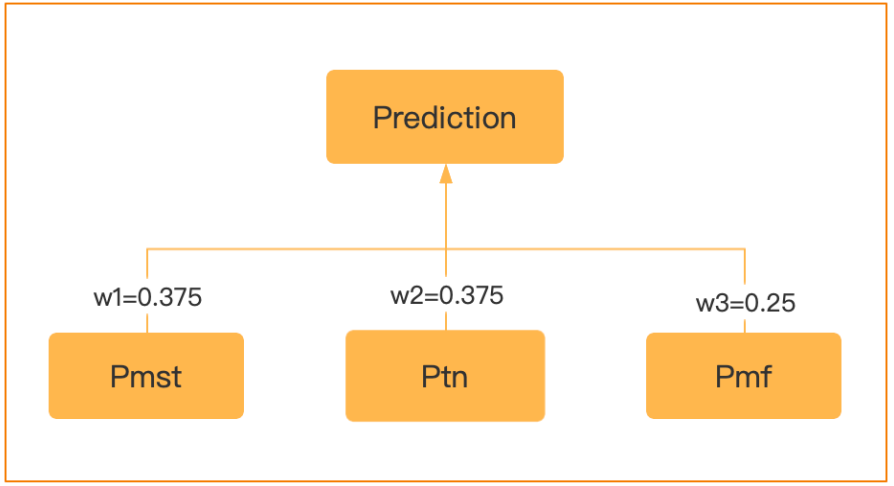
- The model consists of three modules:
  - Multi-Step Tree Module
  - Top-N Module
  - Multivariant Fusion Module



Sieyuan

Method - Ensemble

- Hybrid forecasting methods combine the ability of multiple models to better accommodate changes.
- In the weighting method, we assign a weight coefficient to the prediction of each model according to the effectiveness of the model, and fine tune the weight through experiments to achieve better results.



Result

Table 1: Comparison of Schemes

Score	Models	Schemes
-45.15699	LightGBM + GRU	A
-45.12303	LightGBM + GRU	B
-45.10273	LightGBM + GRU + Local-Ensemble	C
-44.92340	LightGBM + GRU + Local-Ensemble	D

- (A) LightGBM + GRU, GRU uses top 10 turbines models to predict, and the results of LightGBM and GRU rectified by one step
- (B) LightGBM + GRU, GRU uses top 10 turbine models to predict, and rectifies the results by multi-step, at the same time, the results of LightGBM rectified by one step
- (C) LightGBM + GRU + Local-Ensemble, based on the previous one, Local-Ensemble model is incorporated
- (D) LightGBM + GRU + Local-Ensemble, the difference from previous ones, is that GRU use top 5 turbine models to predict, and the results of LightGBM and GRU both rectified by multi-step

Conclusion

- In the future, we will continue to improve our solution, integrate multiple models into a single model, reduce the number of models, simplify the complexity of training and practical applications.
- And continue to explore the coefficient adjustment of multi-model mixed prediction, and explore more effective time series problem processing methods.

Sieyuan

THANK YOU

trymore: Solution to Spatial Dynamic Wind Power Forecasting for KDD Cup 2022

Hanhan Liu
Sieyuan Electric Co., Ltd
Shanghai, China
zhaizhixiang@yahoo.com