Spatial dynamic wind power forecasting using lightGBM and multi-variate LSTM with hierarchical coherence constraints

Team: Dynamo
Ai Hongfeng
Wu Wenqi
Zhang Chaodong

Oppo Research Institute

Contents

1. Data Preprocessing & Featuring Engineering
2. lightGBM Model
3. Multi-Variate LSTM
4. Ensemble

Anomalies Preprocess

Nearly **30%** anomalies

- Extreme values: time series constrained to a narrow region after standardization.
- Anomalies: cause difficulty for model learning temporal dependencies.

data	Anomalies	Abormal Percentage	Possible Causes	Preprocessing Methods
Wspd	<0	0	-	-
Wdir	out of [-180,180]	0.002%	Occasional Measurement Error	Interpolation
Ndir	out of [-170,170]	0.3%	Occasional Measurement Error	Interpolation
Wind direction (Wdir+Ndir)	nearly random distribution	-	Zero Angle Difference	Wind Angle Adjustment
Etmp	out of $[\mu_t - 3\sigma_t, \mu_t + 3\sigma_t:]$	6%	Broken Thermometer	Median Filling
Pab1/2/3	>89	20.9%	Power Scheduling	Drop
Patv	<0 Wspd>2.5 and Patv<0 Pab1/2/3>89 Wdir is out of [-180,180] Ndir is out of [-170,170]	29%	Power Scheduling Turbine Renovation Measurement Error	Treat <0 as 0 LightGBM Wind Speed – Max Power Curve Adjustment

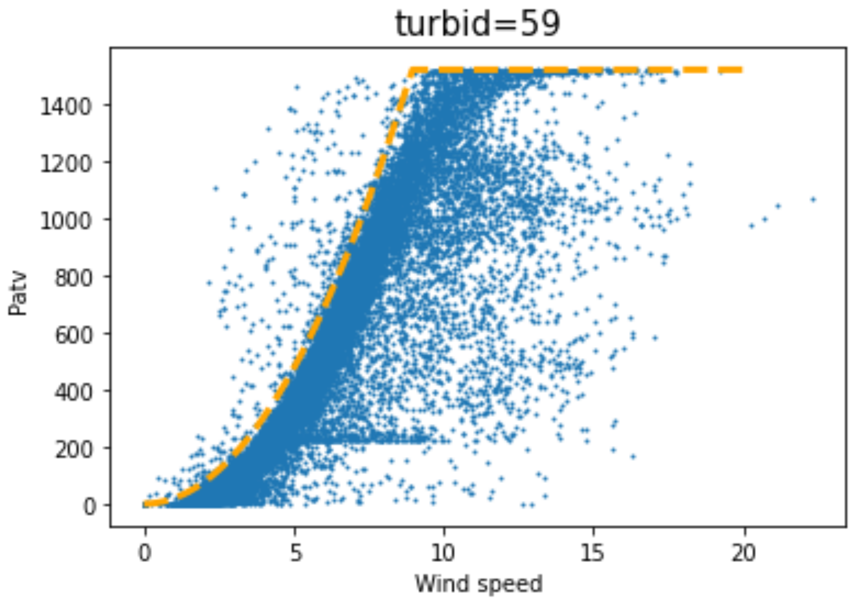
Anomalies Preprocess : Wind Speed - Max Power Curve

Assumption: For a fixed wind speed v, there is a maxinum power $P_{max.}$ Other reasons cause the actual power to be blow $P_{max.}$

$$P_{max} = f(v, \theta) = f(v; \alpha, \beta, s) = \begin{cases} \beta v^{\alpha}, v < s \\ \beta s^{\alpha}, v \geq s \end{cases}$$

hitograms of wind speed, center of group g is v_g

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \left[C \sum_{(i,g), i \in g} (P_i - f(v_g; \theta))^+ + \sum_g (f(v_g; \theta) - \max_{i \in g, P_i \leq f(v_g;\theta)} P_i \ ) \right]$$



Hyperparameter C decides the penalty above the curve.
Solved by python *scipy* package directly.

- Sparse solution: max curve is decided by data above or just below the curve .
- Data above the curve was treated as anomalies and adjusted to the corresponding value on the curve.

Anomalies Preprocess : Wind Angle Adjustment

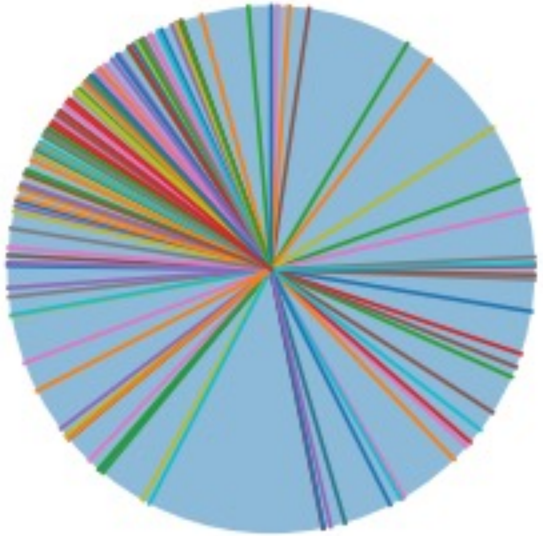
Wind angle respect to north(zero angle) = Ndir + Wdir?

Nearly randomly distributed, might be caused by zero angle shift.

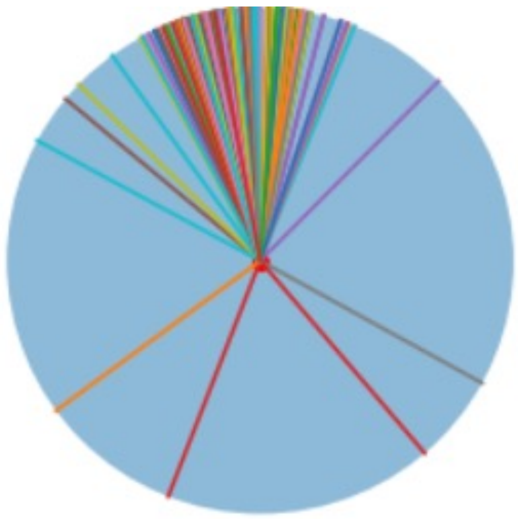
similar directions except a few outliers

Assumption: When wind speed >10m/s, wind angles at all turbines are the same.

wind directions



Before adjustment



After adjustment

Feature Engineering

Features used for our methods are summarized as follows:

1. Circular Feature:

- Wind Direction: $sin\frac{2\pi w_t}{360}$, $cos\frac{2\pi w_t}{360}$.
- Minutes of the day: $\sin\frac{2\pi m_t}{1440}$, $cos\frac{2\pi m_t}{1440}$.

2. Max Power Features: $P_{max} - P_{gt}$.

3. Sliding Window Features:

- Features: Patv、 Wspd、Itmp、Prtv、 Max Patv and Wdir.
- Aggregation: mean, maximum, minimum, variance, and median.

4. tsfresh Features

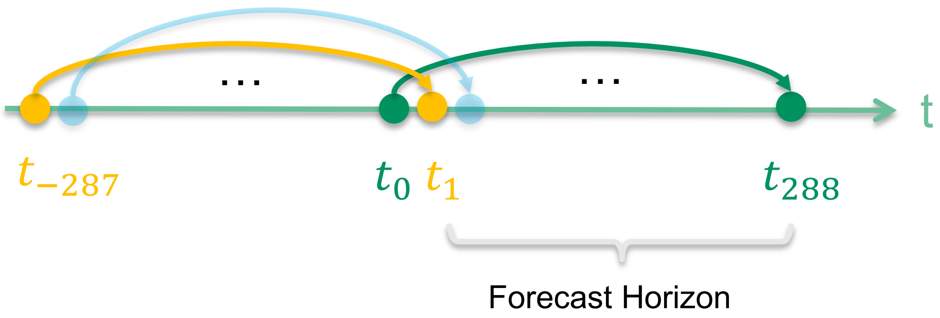
Top 59 features were generated included fourier entropy and coefficient, auto-correlation coefficient, etc.

LightGBM

Since the data distribution is unknown, LightGBM with better generalization performance is more preferable than XGBoost.

LightGBM

Forecast the 288th value ahead



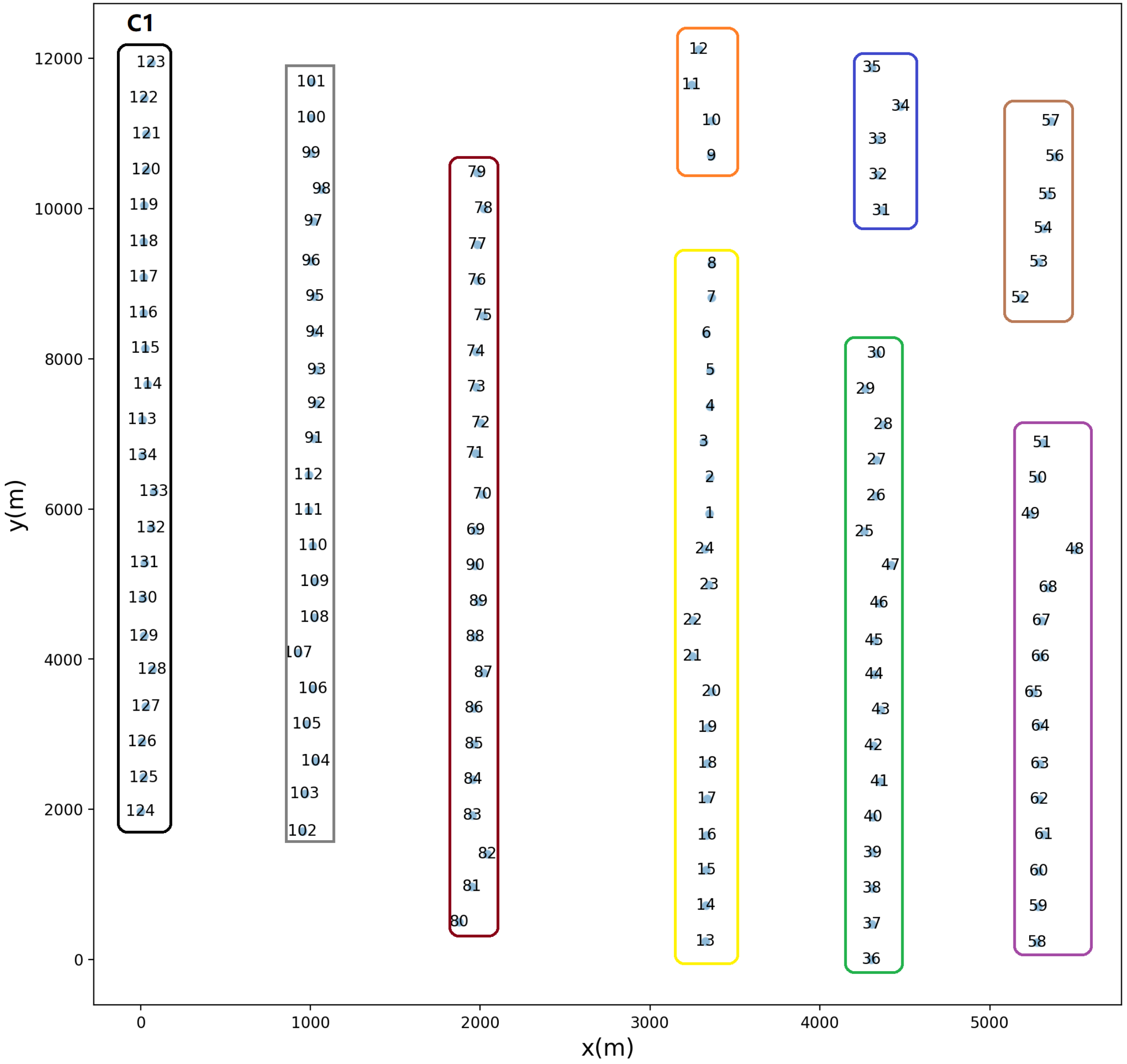
Model Configuration (successive)	Offline MAE	Offline RMSE	Offline Score	Online Score
Baseline	39.37	42.83	41.10	41.92(Phase I)
Filling Unknown	41.85	44.92	43.39	41.02(Phase I) 0.9↓
Max Power-Wind Speed	41.31	44.42	42.87 0.52↓	46.57(Phase III)
tsfresh Features	43.04	45.99	44.51	46.54(Phase III) 0.03↓

Note: Lower Score represents better performance.

Conclusion:

- Abnormal Preprocessing significantly reduced the online score in the Phase I.
- Adding max power-wind speed adjustment and features somehow reduced offline error.
- More features engineering is limited to the 288th-forecasting task.

Hierarchical Coherence Constraints



Turbines are clustered into groups

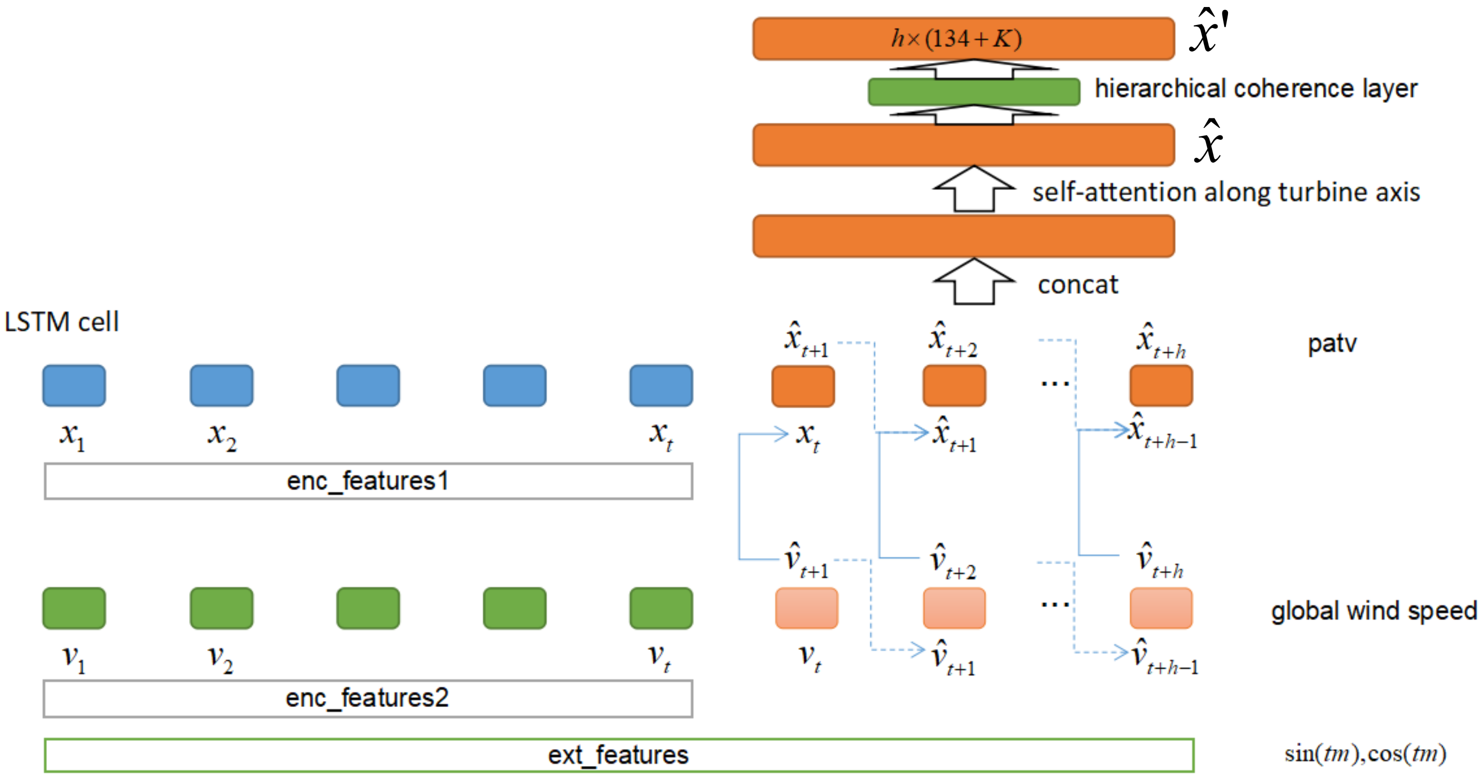
$$x_{c1} = \sum_{i=113}^{134} x_i$$

x is the power of all groups and turbines
All constraints can be writen as

$$Ax = 0$$

Matrix A is solely decided by the hierarchical structure of groups

Multi-Variate LSTM with Hierarchical Coherence Constraints



multi-variate: predict the power of all turbines and groups $\hat{x}$

hierarchical coherence layer: satisfy constraints with minimal change:

$$\hat{x}' = \underset{\hat{x}'}{\operatorname{argmin}} \frac{1}{2} ||\hat{x}' - \hat{x}||$$

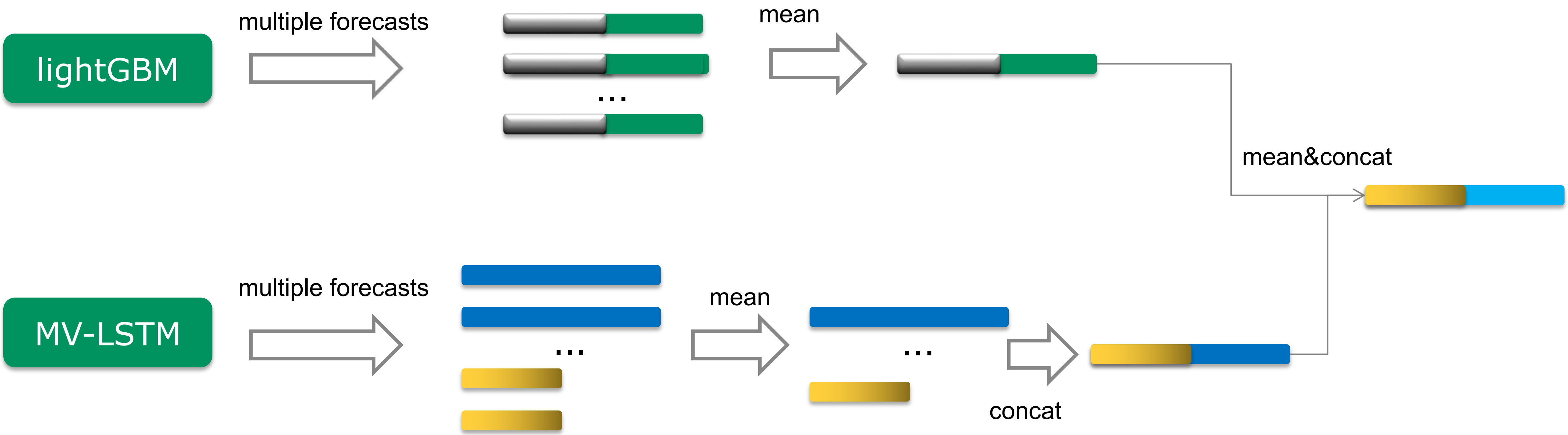
s.t.

$$A\hat{x}' = 0$$

closed form solution:

$$\hat{x}' = (I - A^T(AA^T)^{-1}A)x$$

Ensemble



	lightGBM	MV-LSTM	ensemble
offline	40.88	42.95	41.2
online	46.73	45.87	45.64

Q&A