Complementary Fusion of Deep Spatio-Temporal Network and Tree Model for Wind Power Forecasting (Team:HIK)

Linsen Li[1†,#], Qichen Sun[2†], Dongdong Geng[3#], Chunfei Jian[4#], Dongen Wu[5†],
Shiliang Pu[6#]
[†]*College of Computer Science, Zhejiang University, China*
[#]*Hikvision Research Institute, China*
{[2]sunqichen, [5]wude}@zju.edu.cn
{[1]lilinsen, [3]gengdongdong, [4]jianchunfei , [6]pushiliang.hri}@hikvision.com

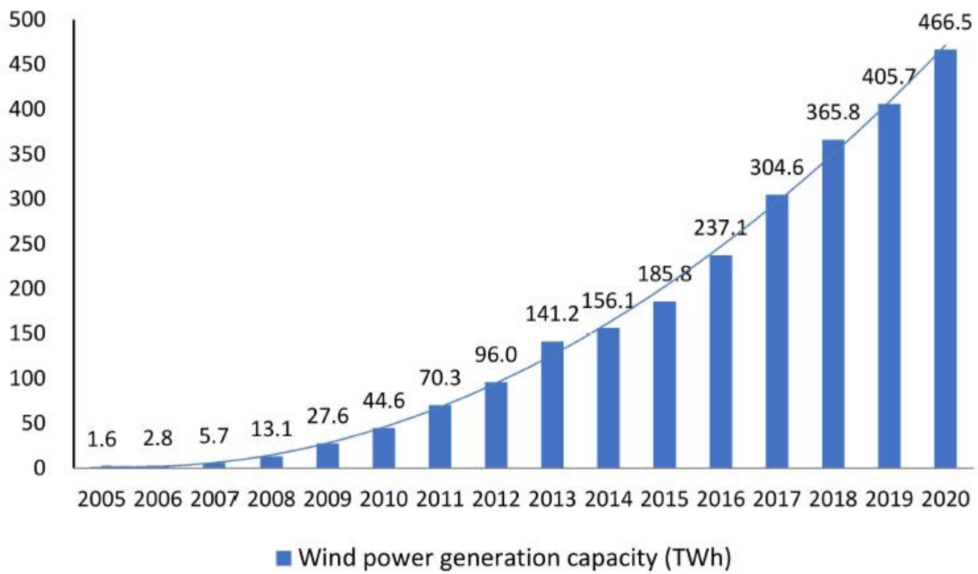
Outline

- **Motivation**
- Related Work
- Preliminaries
- Methods
- Experimental Evaluation
- Conclusions

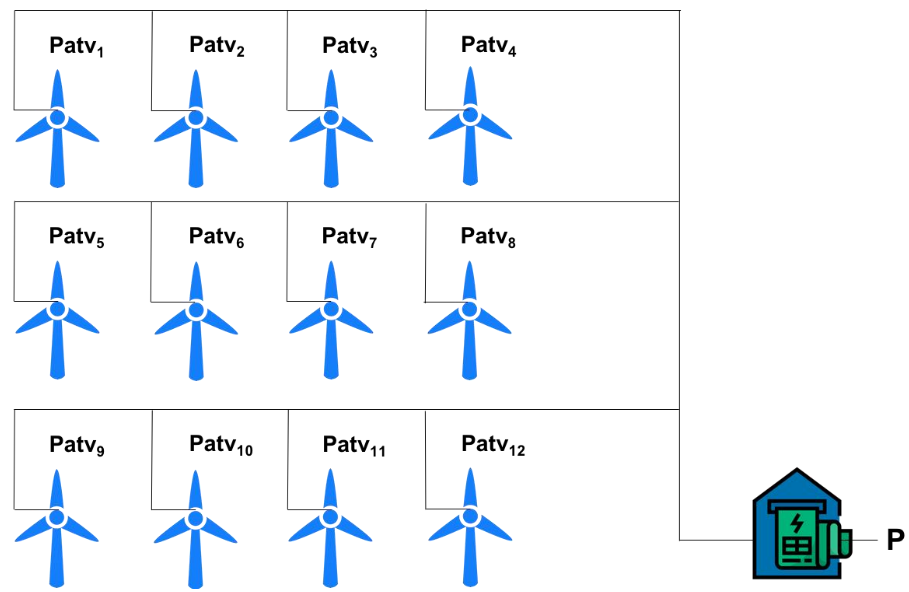
Motivation

Important & Necessary

- Wind energy plays an important role in energy conservation and emission reduction
- Wind Power Forecasting(WPF) has been recognized as one of the most critical issues in wind power integration and operation



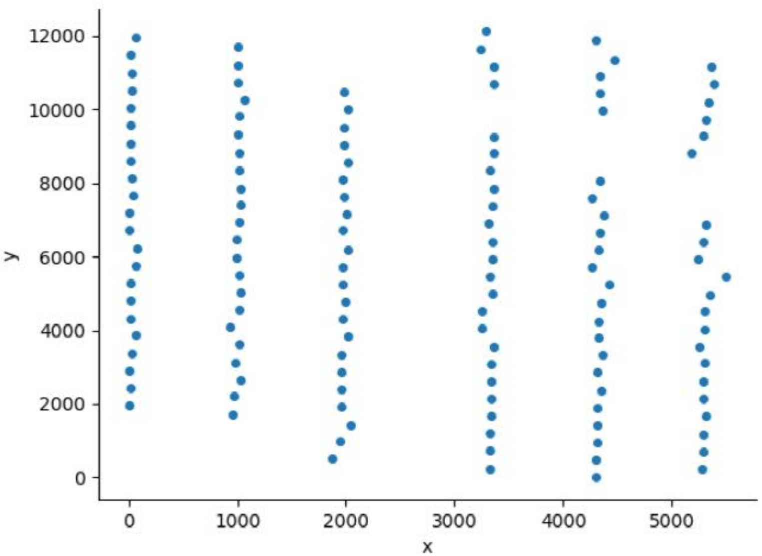
Wind power generation capacity trend in China
(Data source: National Bureau of Statistics)



Example of Wind power forecasting

Motivation(Cont.)

- KDD CUP 2022: Spatial Dynamic Wind Power Forecasting
  - Spatial distribution: the relative location of all wind turbines
  - Dynamic context: important situations and turbine internal contexts



Spatial distribution of all wind turbines

- Objective
  - Predict power generation of 134 wind turbine ahead of 48 hours
  - Average of RMSE and MAE is used as the main evaluation
  - Random sampled stride time steps to evaluate the submitted models
- Requirements
  - No external data is allowed to use
  - Submitted file should not exceed 200MB
  - Maximum length of input time series is 14 days

Outline

- Motivation
- **Related Work**
- Preliminaries
- Methods
- Experimental Evaluation
- Conclusions

Related Work

Methods

Methods	Type	Techniques
AutoFormer [Wu et al.]	Time Series Model-based	Decomposition， Auto-Correlation， Transformer
AGCRN [Lei et al.]	Spatio-Temporal Model-based	Node Adaptive Learning, Data Adaptive Graph Generation
QC-WPF [Browell et al.]	Ensemble Model-based	GBDT, Boosted Generalized Additive, Quantile Combination

Limitations

- A single or no relation between turbines(e.g., Euclidean distance) to construct the adjacency graph, which ignores the complicated relation between nodes
- Most ensemble strategies fuse the models statically

Outline

- Motivation
- Related Work
- **Preliminaries**
- Methods
- Experimental Evaluation
- Conclusions

Preliminaries

Spatial Dynamic Wind Power Forecasting

- **Turbine Time Series.** $X = \{x_1, x_2, \ldots, x_N\} \in R^{N \times T \times D}$ represents multivariate time series generated by all turbines, where N is the number of wind turbines, $x_i \in R^{T \times D}$ is the status of turbine i at timestamp t.

- **Spatial Correlation Graph.** $G = (V, E)$ is used to capture the spatial correlation, where $V = \{v_i\}_{i=1}^{N} (|V| = N)$ is a set of vertices and E is the edge set.

- **Spatial Dynamic Wind Power Forecasting** is formulated as learning a function $\mathcal{F}_\theta$ to forecast next τ steps data based on the past T steps historical data:

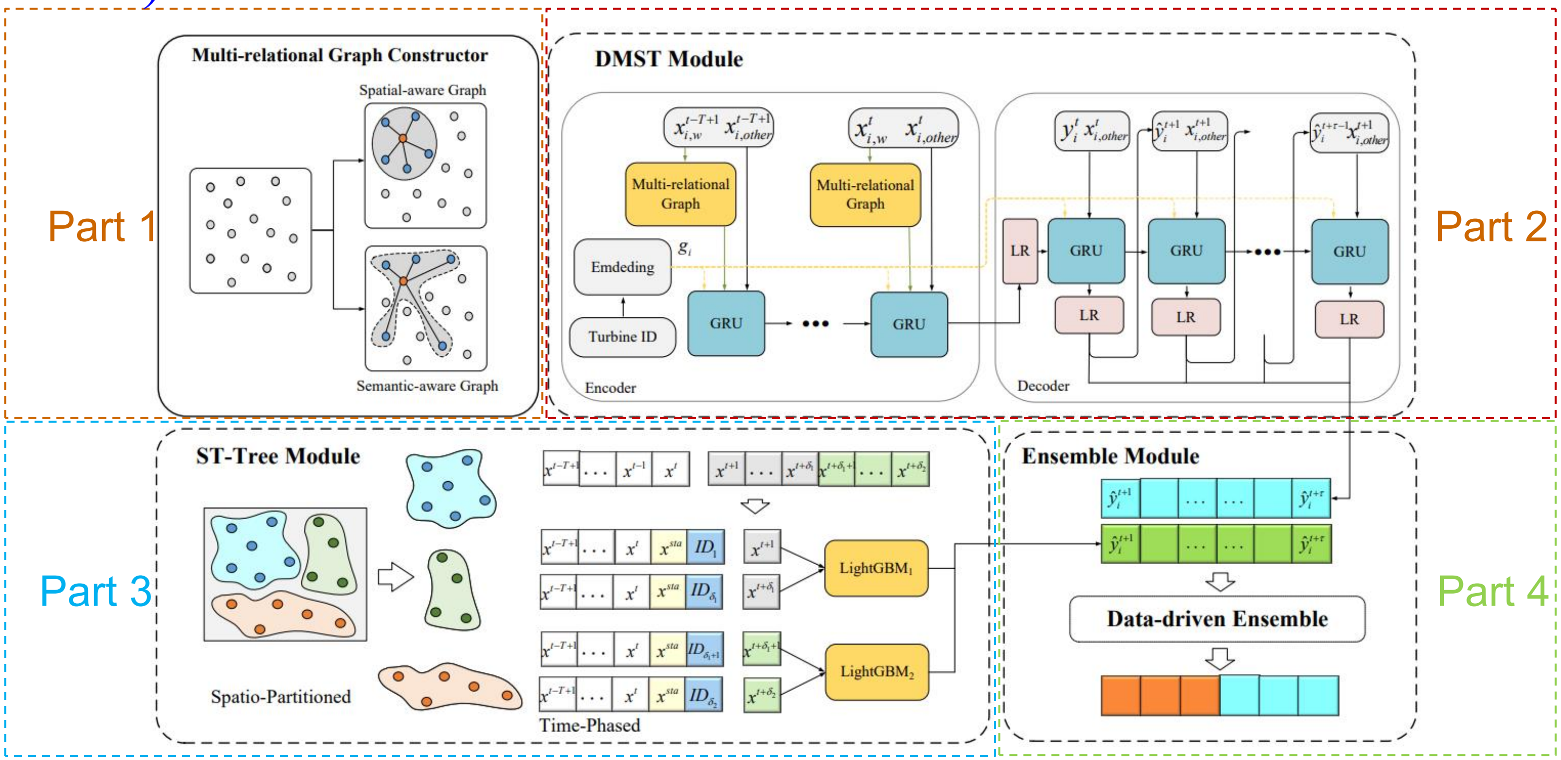
$$\left\{\hat{Y}^{t+1}, \hat{Y}^{t+2}, \cdots, \hat{Y}^{t+\tau}\right\} = \mathcal{F}_\theta\left(X^t, X^{t-1}, \cdots, X_{t-T+1}; \mathcal{G}\right)$$

Outline

- Motivation
- Related Work
- Preliminaries
- **Methods**
- Experimental Evaluation
- Conclusions

Overview

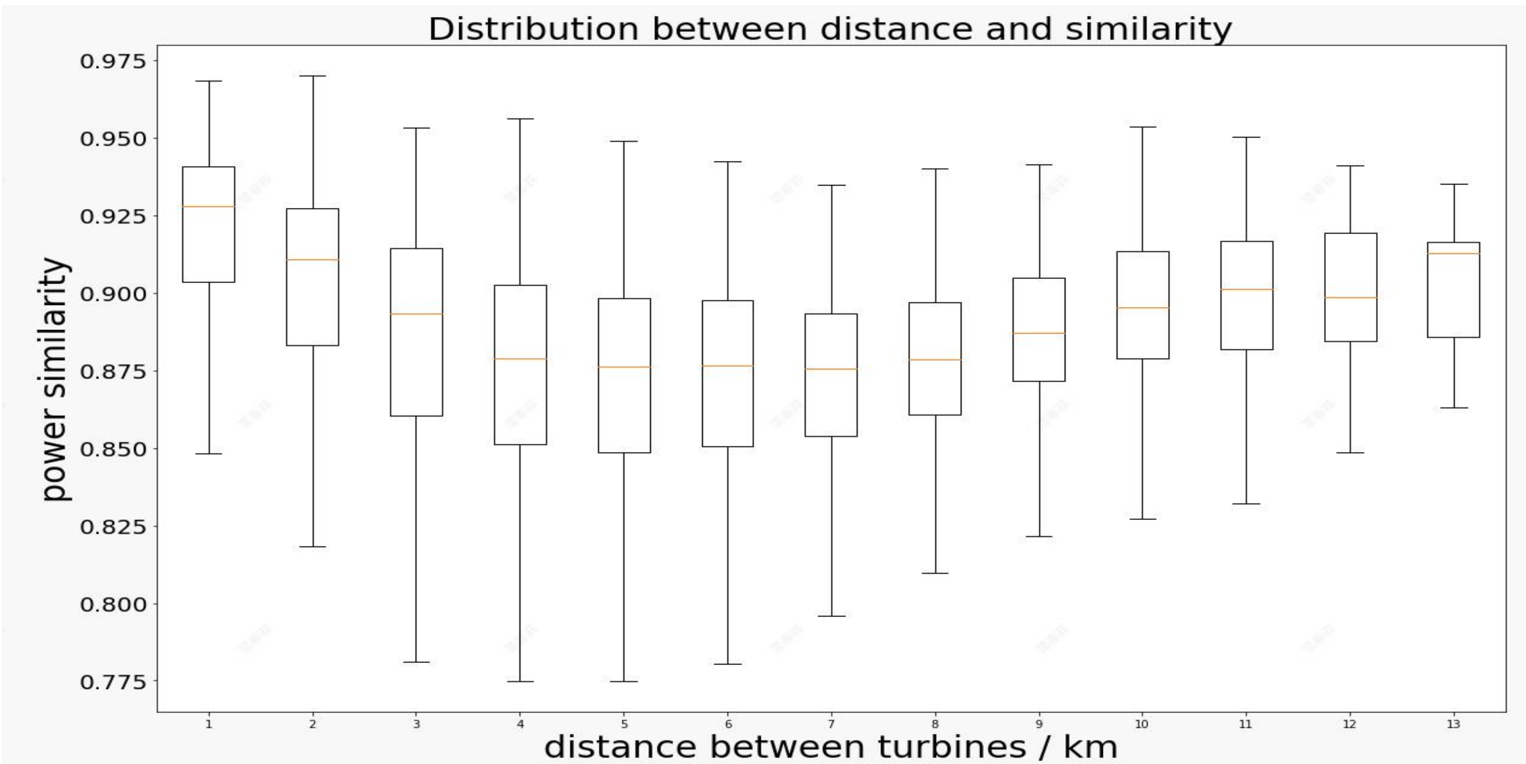
- Complementary Fusion of Deep Spatio-Temporal Network and Tree (FDSTT) Model



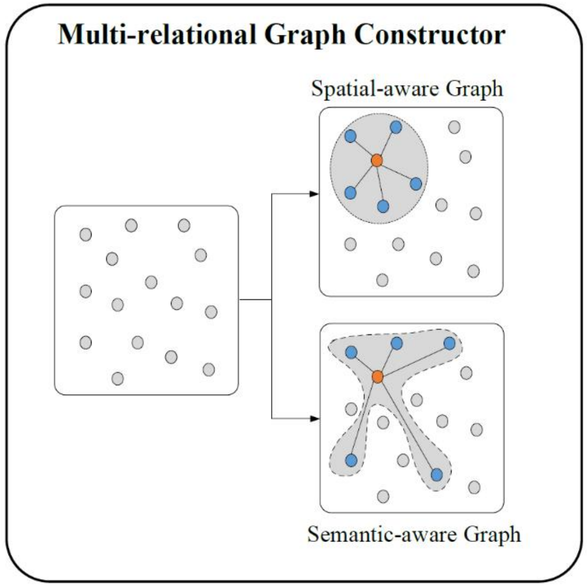
- Part 1: Multi-relational Graph Constructor
- Part 2: Deep Multi-relational Spatio-Temporal (DMST) Module
- Part 3: Spatio-partitioned and Time-phased Tree (ST-Tree) Module
- Part 4: Ensemble Module

Part 1: Multi-relational Graph Constructor

- Correlation between power similarity and distance between nodes
  - Distance is an important factor to similarity between nodes
  - High similarity between nodes still happens when distance is large.



Part 1: Multi-relational Graph Constructor(Cont.)



Spatial-aware Graph

- Calculate the Euclidean distance
- Take the top-K nearest nodes as the neighbors of node

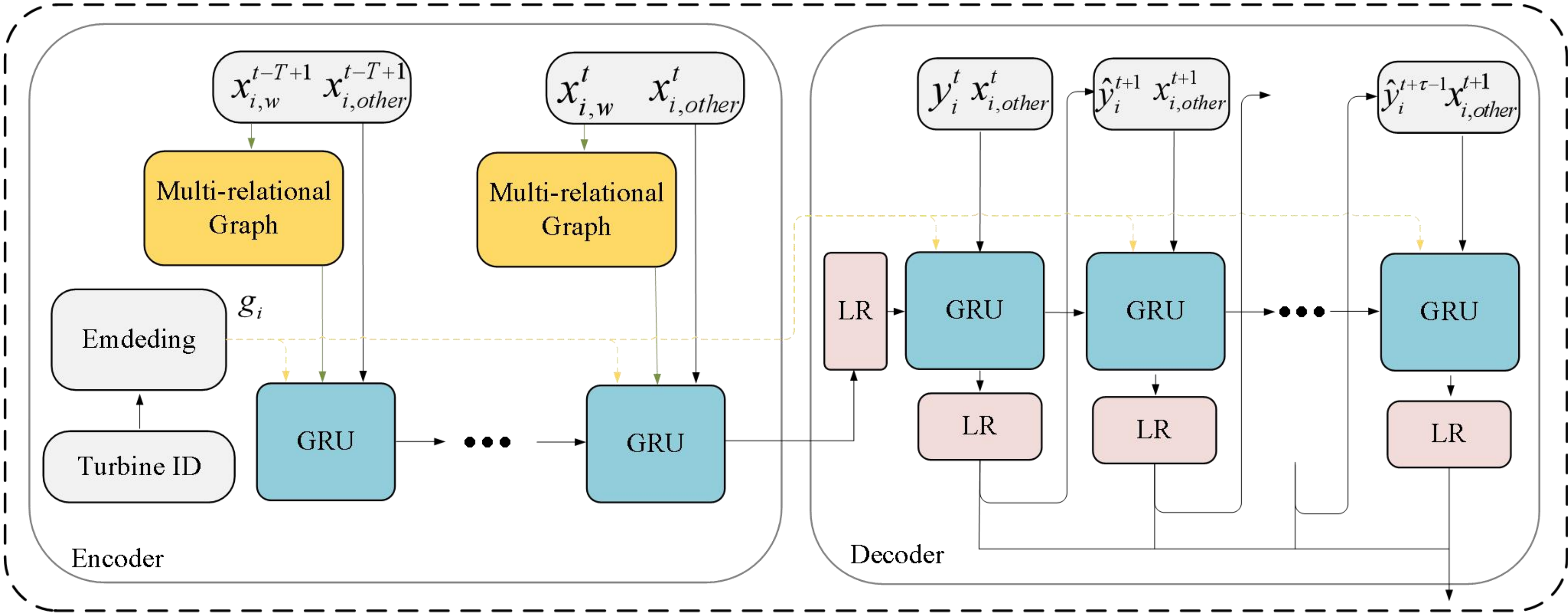
$$A(i,j) = \begin{cases} 1, & j \in N(i) \\ 0, & j \notin N(i) \end{cases}$$

Semantic-aware Graph

- Calculate the differential similarity
- Obtain the top-K most similar nodes as semantic neighbors for each node

$$Sim(i,j) = \sum_{t=1}^{T} \left((x_{i,w}^{t} - x_{i,w}^{t-1}) \cdot (x_{j,w}^{t} - x_{j,w}^{t-1}) \right)$$

Part 2: Deep Multi-relational Spatio-Temporal Network



Features:

- Spatial features:

  $\hat{x}^t_{i,w}$=CONCAT($x^t_{i,w}$, AGGR($x^t_{u,w}, u \in N(i)$))

- Turbine embedding: g_i
- Temporal features: $\hat{x}^t_{i,w}$, $x^t_{i,\, other}$

Model:

- Encoder

$$x_i^t = \text{CONCAT}(g_i, \hat{x}^t_{i,w}, x^t_{i,other})$$
$$h_i^t = \text{GRU}(x_i^t, h_i^{t-1})$$

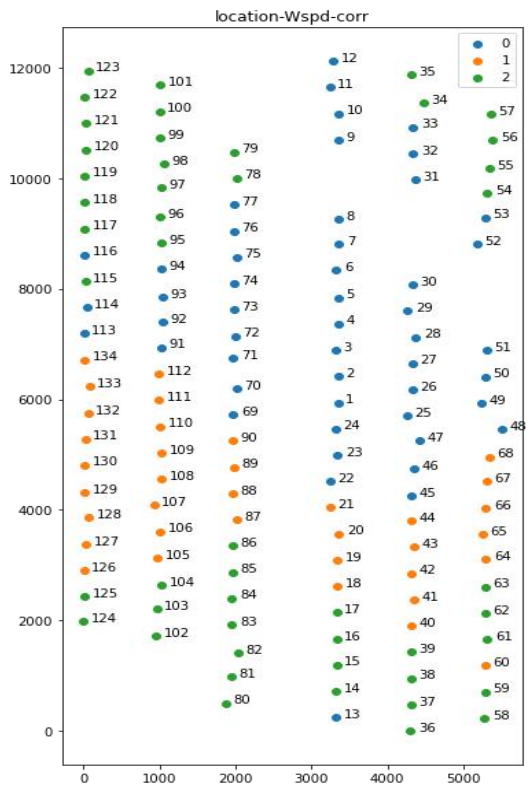
- Decoder

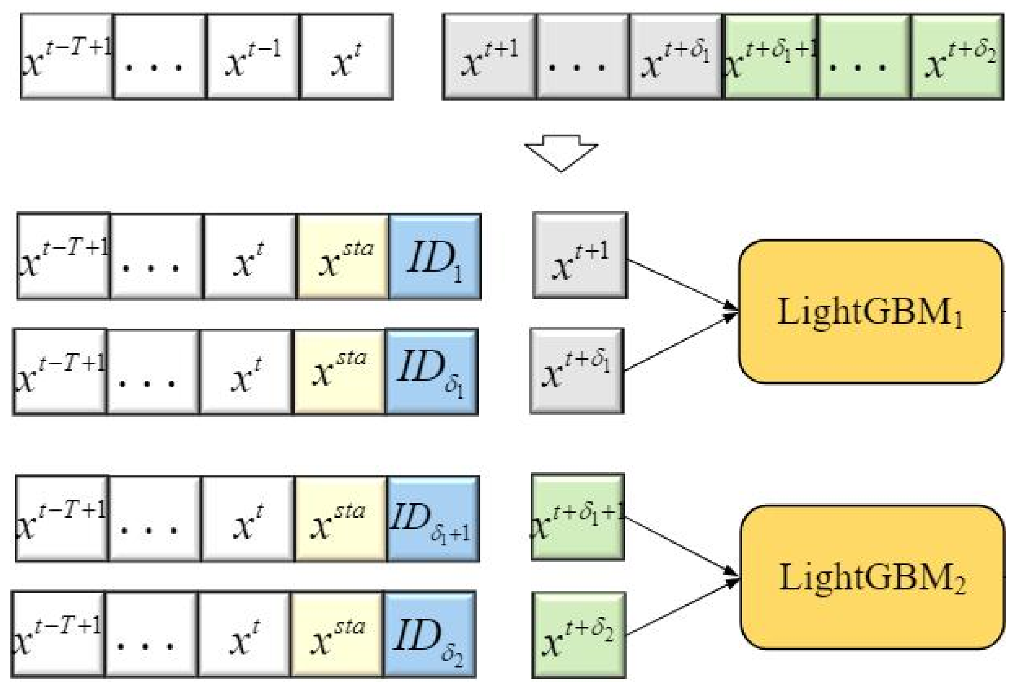
$$x_i^{t+1} = \text{CONCAT}(g_i, \hat{y}_i^{t+1}, x^{t+1}_{i,other})$$
$$h_i^{t+1} = \text{GRU}(x_i^{t+1}, h_i^{t+1})$$
$$\hat{y}_i^{t+2} = \text{Linear}(h_i^{t+1})$$

Part 3: Spatio-Partitioned Time-Phased Tree Model

◻ Spatio-Partitioned Time-Phased

- Cluster wind turbines by Pearson coefficient correlation for several spatial partitions.
- Separate timestamp into several time phases to reduce model numbers
- Build a tree model for each spatio-partitioned time-phased pattern.





Part 4: Ensemble Model

Data-driven ensemble strategy

- Adjust the ensemble architecture to fit input distribution dynamically.

Algorithm 1: Data-driven Ensemble Strategy

Input: $X_p \in \mathbb{R}^{N \times T \times 1}$ is the power value of model input,
$\hat{Y}_1 \in \mathbb{R}^{N \times \tau \times 1}$ is the predicted value of DMST,
$\hat{Y}_2 \in \mathbb{R}^{N \times \tau \times 1}$ is the predicted value of tree model, ϕ is the baseline value, Δ_{low} and Δ_{up} are the judgment threshold of the current power level, β is timestamp.

Output: $\hat{Y}$ is the predicted value

```
Initialize s = Mean(X_p[:, -5 :, :])
Initialize Ŷ = 0
if s < Δ_low or s > Δ_up then
    Ŷ[:, : β, :] = Ŷ_2[:, : β, :]
else
    Ŷ[:, : β, :] = 0.5 × Ŷ_2[:, : β, :] + 0.5 × Ŷ_1[:, : β, :]
Ŷ[:, β :, :] = Ŷ_1[:, β :, :] + φ
```

Outline

- Motivation
- Related Work
- Preliminaries
- Methods
- **Experimental Evaluation**
- Conclusions

Experimental Evaluation

Datasets

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

Competitors

- **Time Series Model:** GRU, AutoFormer, SCINet
- **Spatio-temporal Model:** AGCRN, GWNET, DCRNN, ASTGCN, DST

Evaluation Metrics

$$s_{t_0}^{i} = \frac{1}{2}\left(\sqrt{\frac{\sum_{j=1}^{288}(Patv_{t_0+j}^{i} - \overline{Patv}_{t_0+j}^{i})^2}{288}} + \frac{\sum_{j=1}^{288}|Patv_{t_0+j}^{i} - \overline{Patv}_{t_0+j}^{i}|}{288}\right)$$

Experimental Evaluation (Cont.)

Performance on Official KDD CUP 2022 Test Dataset

- Win the **1st** in the final Phase and the **2nd** in Phase 2

Table 1: Online scores with different models. The footnotes (e.g., 1st) in the table denote the online rank of the corresponding method in Baidu KDDCup 2022.

Method	Phase 1	Phase 2	Phase 3
AutoFormer	45.5570	–	–
SCINet	46.4679	–	–
AGCRN	41.3100	–	–
GRU (Baseline)	42.3019	46.9968	–
ST-Tree	**40.7903** (3rd)	45.1745	–
GWNET	–	48.8300	–
DCRNN	–	47.3043	–
ASTGCN	–	48.0889	–
DST	–	44.4205	–
DMST	–	44.2845	–
FDSTT (w/o avg)	–	44.0942	45.0405
FDSTT (w/o ϕ)	–	44.0732	45.0169
FDSTT	–	**44.0536** (2nd)	**44.9171** (1st)

Experimental Evaluation (Cont.)

Performance on Self-constructed Test Dataset

Table 2: The performance of different methods.

Method	MAE	RMSE	Score
GRU (baseline)	37.0174	47.0442	42.0308
AutoFormer	42.8972	54.3160	48.6066
SCINet	40.2794	47.8892	44.0843
GWNET	43.1812	55.1324	49.1568
AGCRN	40.9344	52.4761	46.7052
DCRNN	39.6786	48.4222	44.0504
ASTGCN	37.9236	46.2108	42.0672
DST	36.8172	46.1150	41.4661
FDSTT (ours)	**36.3872**	**45.5246**	**40.9559**

Ablation study

Table 3: Ablation study of different modules of FDSTT .

Method	MAE	RMSE	Score
DMST	36.8119	46.0136	41.4128
w/o mg	36.8172	46.1150	41.4661
ST-Tree	37.0405	46.3101	41.6753
w/o sp	37.3759	46.3857	41.8808
Ensemble	36.3872	45.5246	40.9559
w/o ϕ	36.4568	45.6260	41.0414
w/o avg	36.3410	45.8520	41.0965

Outline

- Motivation
- Related Work
- Preliminaries
- Methods
- Experimental Evaluation
- **Conclusions**

Conclusions

- Propose an ensemble framework for long-term wind power forecasting called FDSTT.
- Present a multi-relational graph constructor to capture the multi-relational dependencies among wind turbines.
- Design a data-driven ensemble strategy.
- Win the 1st place in the KDD Cup 2022.

Thank you !

Questions?

Questions?

lilinsen@hikvision.com