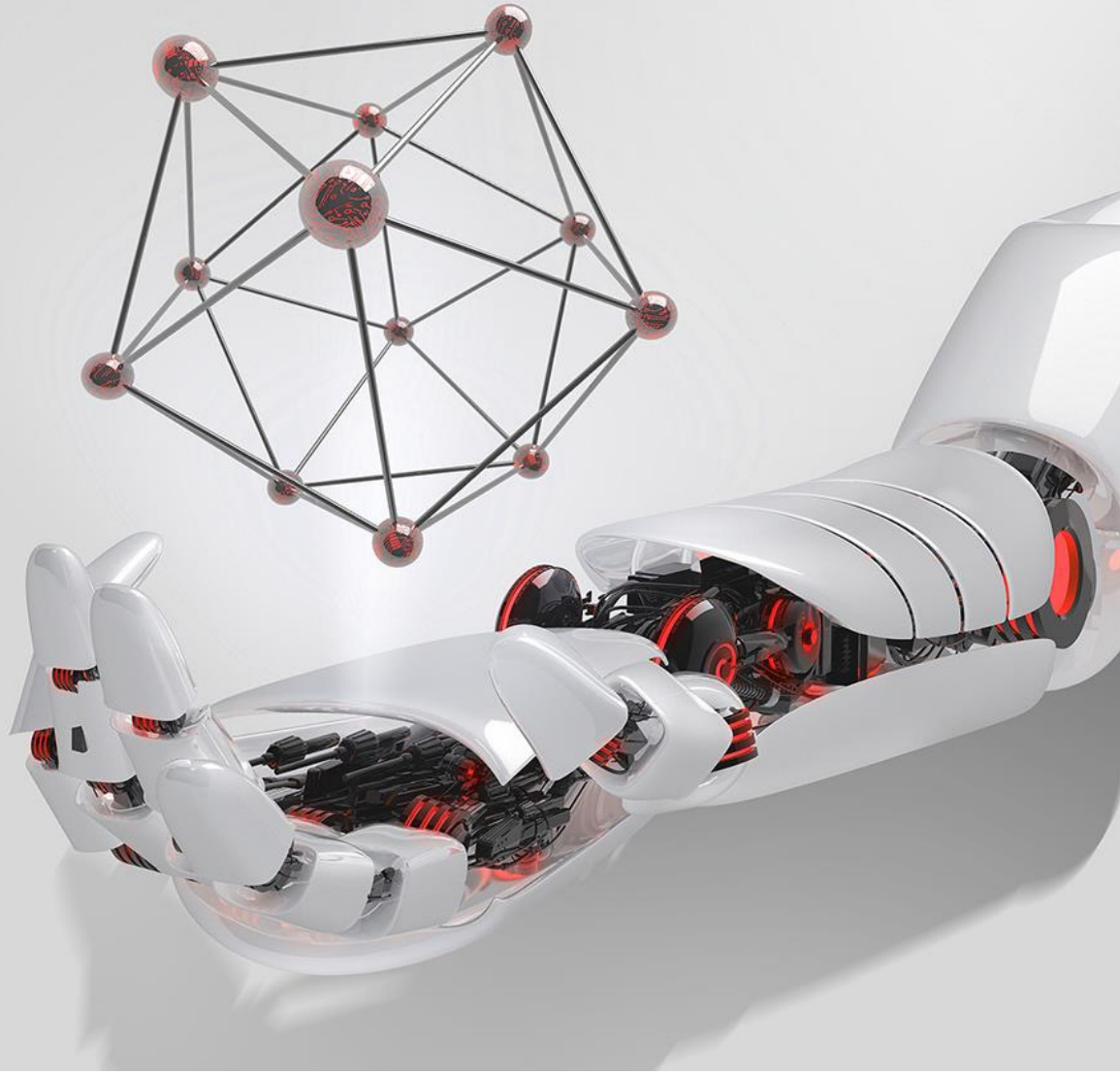
KDD Cup 2022
Spatial Dynamic Wind Power Forecasting Competition




Marcus Kalander


Zhongwen Rao


Chengzhi Zhang

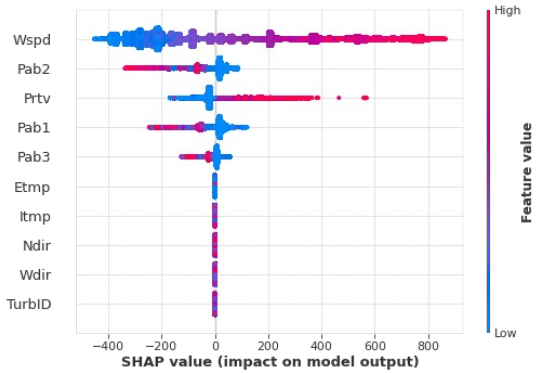
Team: didadida_hualahuala
Date: 2022-08-17
Rank: 6th (regular track).

STATE STREET.

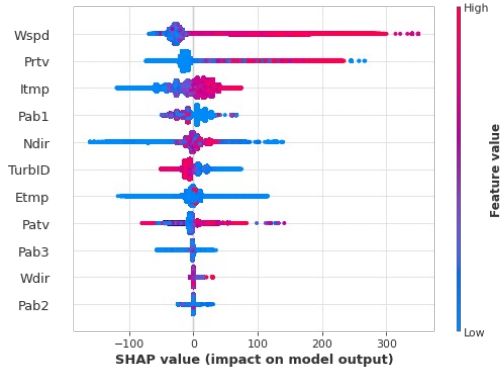
NOAH'S ARK LAB

HUAWEI

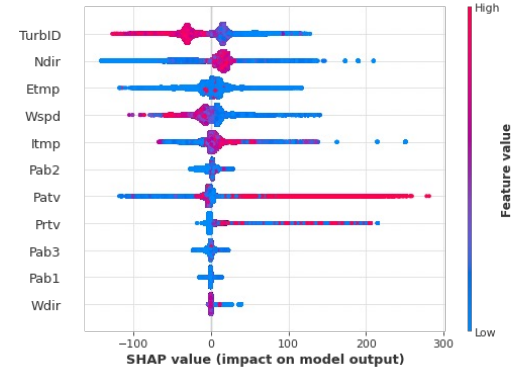
Preprocessing & Feature Engineering



Immediate forecast



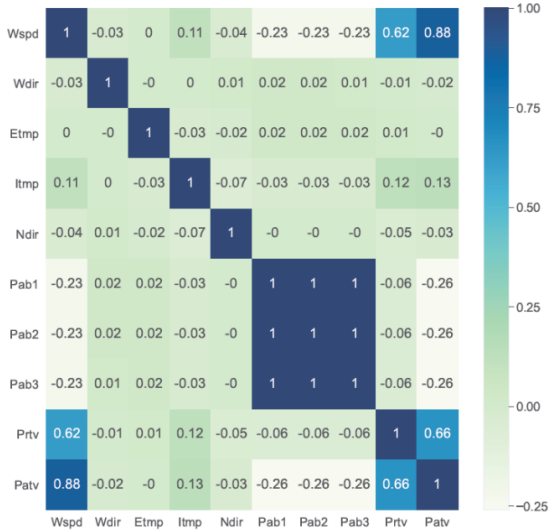
One-day future forecast



Two-day future forecast

- Not all features are important.
  - We remove the directional features $Wdir$ and $Ndir$.
  - We remove the temperature features $Etmp$ and $Itmp$.
- The pitch angles ($Pab1$, $Pab2$, $Pab3$) are perfectly correlated.
  - We merge these to $Pab\_max = \max(Pab1, Pab2, Pab3)$.

Final feature set: {$TrubID$, $Wspd$, Pab_max, $Prtv$, $Patv$}.



HUAWEI

Solution Overview

Fusion of two very different models.

- Modified DLinear (MDLinear): An altered version of DLinear [1].
- Extreme Temporal Gated Network (XTGN): based on stacking gated temporal convolutional networks (TCNs) [2,3] and nearest neighbor information diffusion.

Both models uses a masked loss function that ignores any missing, unknown or abnormal values.

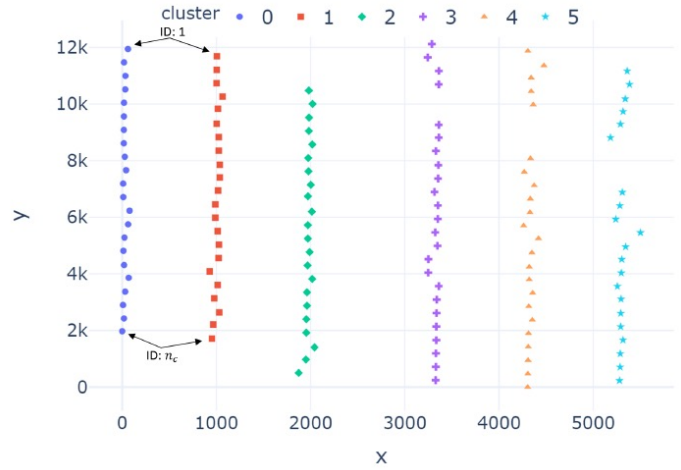
[1]: Ailing Zeng, Muxi Chen, Lei Zhang and Qiang Xu. Are Transformers Effective for Time Series Forecasting?. arXiv, 2022 (May).

[2]: Yann N. Dauphin, Angela Fan, Michael Auli and David Grangier. Language Modeling with Gated Convolutional Networks. arXiv, 2017.

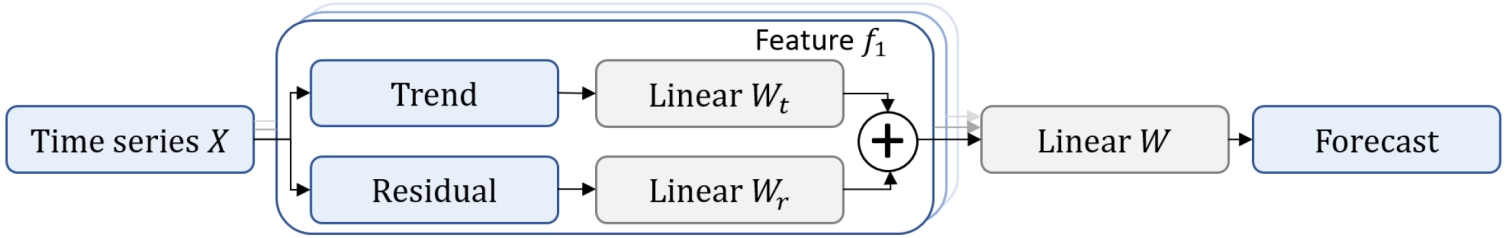
[3]: Colin Lea, Michael D. Flynn, Rene Vidal, Austin Reiter, and Gregory D. Hager. Temporal convolutional networks for action segmentation and detection. CVPR, 2017.

MDLinear: Modified DLinear

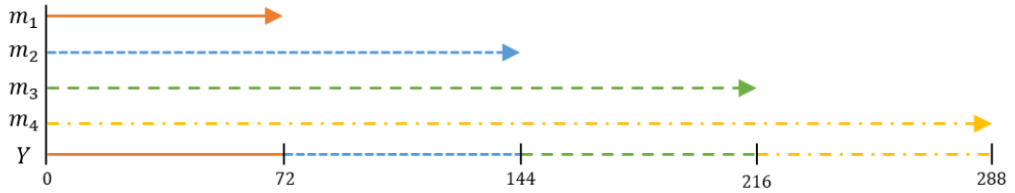
- Additional feature engineering
  - The $TurbID$ is split into 2 separate features: $cluster$ and id.
  - New feature $cluster\_avg$ with average $Patv$ of each cluster.



- Model overview
  - Decompose the time series of each feature into trend and residual.
  - Apply a linear layer on each and merge.
  - Merge the results of each feature to obtain the forecast.



- Multi-horizon training and forecasting:
  - 4 separate models with different forecast horizons [72, 144, 216, 288] are trained and fused.



HUAWEI

XTGN: Extreme Temporal Gated Network

- Uses a gated extension of TCN that are stacked in layers.
- The output of each layer is added together and passed through a convolutional and linear layer.
- During inference, we apply an information diffusion approach to incorporate neighbor information.
  - The $n = 49$ nearest neighbors are considered.
  - Final output for wind turbine x with forecast Y: $\alpha \cdot Y + \frac{1-\alpha}{n} \sum_{v \in N(x)} v$.

Temporal Convolutional Network (TCN)

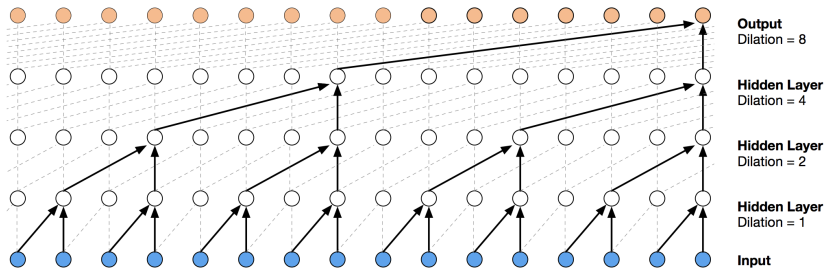
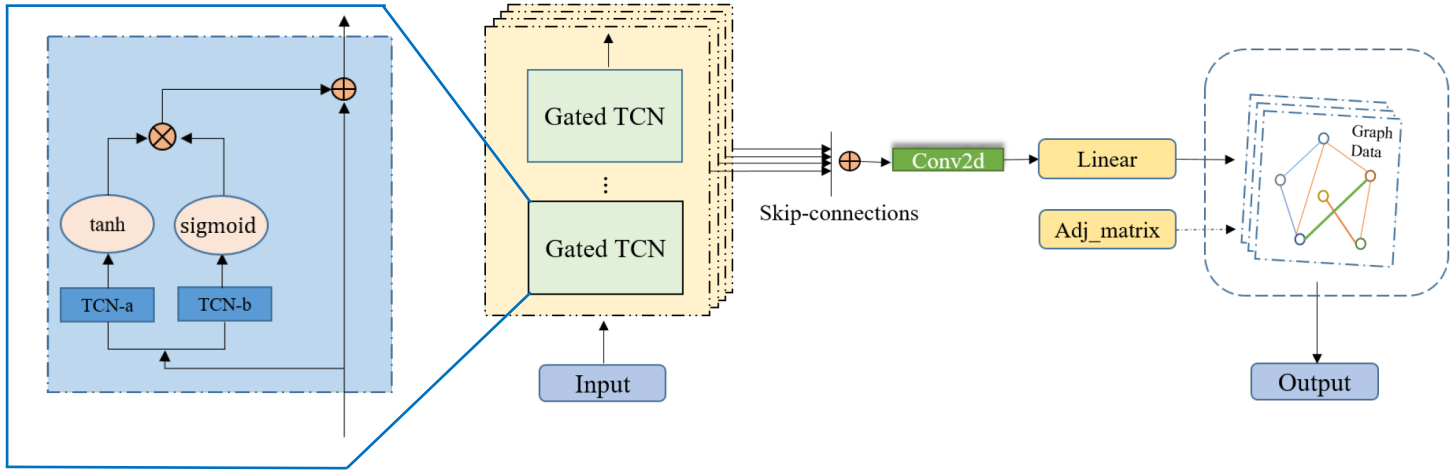


Figure from: https://github.com/philipperemy/keras-tcn

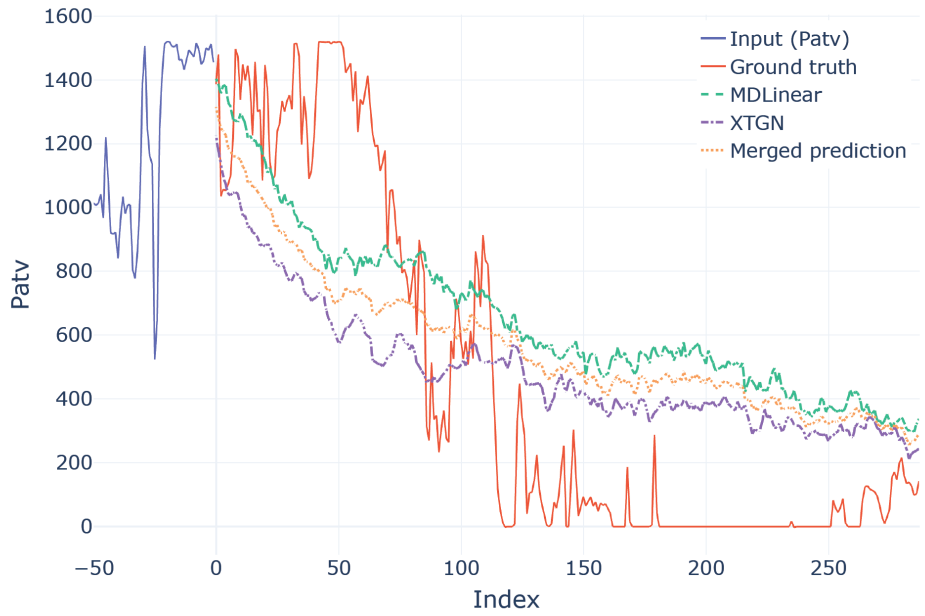


HUAWEI

Fused Model & Results

Simple averaging of the two model forecasts.

$$Y = \frac{Y_{mdlinear} + Y_{xtgn}}{2}$$



Ablation study of different fusion strategies.

Method	RMSE	MAE	Score
Time split 72:216 (MDLinear)	54.24	46.09	50.17
Time split 72:216 (XTGN)	53.74	45.94	49.84
Time split 188:188 (MDLinear)	53.95	45.90	49.92
Time split 188:188 (XTGN)	54.02	46.13	50.07
Average	53.74	45.86	49.80

Offline scores and inference times.

Method	RMSE	MAE	Score	Time (s)
Historical average	56.72	47.86	52.29	121
Moving average	61.56	50.62	56.09	127
GRU	55.13	45.77	50.45	409
GNN	55.39	47.15	51.27	245
LightGBM	53.05	44.89	48.97	4,035
MDLinear (single model)	56.74	48.32	52.53	4,63
MDLinear	53.40	45.53	49.46	1,384
XTGN	54.54	46.50	50.52	227
Fused model	53.74	45.86	49.80	1,420

Summary

Fusion of two considerably different approaches.

- MDLinear
  - Time series decomposition.
  - Exceedingly simple linear network
  - Merged forecast with four models with different forecast horizons.
- XTGN
  - Stacked layers of gated TCNs.
  - Information diffusion to incorporate neighbor forecasts.
- Fused model
  - Simple average proved to give the best result.

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

HUAWEI