NASA AIRATHON

Saisamrit Surbehera & Vishweshwar Tyagi

Data Science Institute
Columbia University

March 22, 2022

Brief Methodology

- Pre-processed data to obtain aggregate statistics and lower dimensional representations using variational auto-encoders
- We only used MAIAC satellite metadata
- Models tested include hiearchical forecasting, feed forward networks, LightGBM and their ensembles
- Also experimented with RNNs, Prophet and TemporalFusionTransformer

Preprocessing

- We followed two separate preprocessing approaches, one based on aggregating feature statistics and other based on obtaining lower dimensional representation of satellite metadata
- For the first apporach, each MAIAC hdf file was mapped to its respective grid-id using an alignment dictionary present in its metadata. Several hdf files belonging to the same grid-id and datetime were summarized using statistics such as mean, min, max and standard deviation. This was done for nine out of the thirteen datasets available in the hdf files.
- For the second approach, we used Beta-VAE to reduce the dimensions from 1200×1200 images to 128×2 vectors reflecting the mean and standard variance. Overall 8 different models were used to for each specific columns. These reflect different parts raster of the data.

Baseline

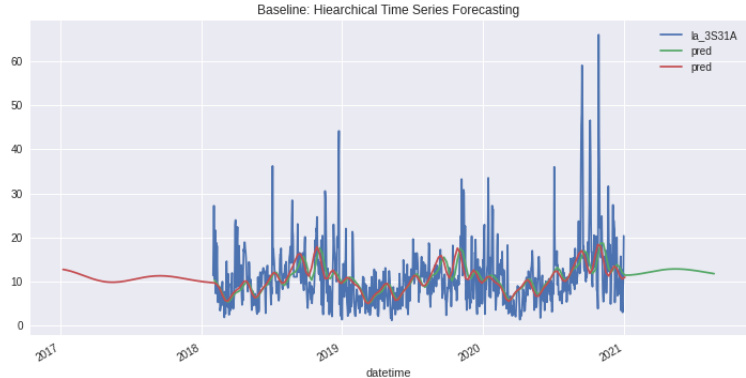
- train-labels as given have an inherent hierarchy, where the cities occupy the upper level and the grid ids occupy the lower level
- We tried incorporating this hierarchy with fb prophet using the scikit-hts framework
- With optimal combination using variance-weighted ordinary least squares as the revision method, we obtained R^2 : 0.5011 and RMSE : 52.78
- We then ensembled it with our feed forward network trained on aggregate features (which scored R^2 : 0.4573) to observe a massive increase in performance, achieving R^2 : 0.5609 and RMSE : 49.52
- We also ensembled the hierarchical model with feed forward network trained on auto-encoded features to obtain R^2 : 0.5758 and RMSE : 48.67

Results

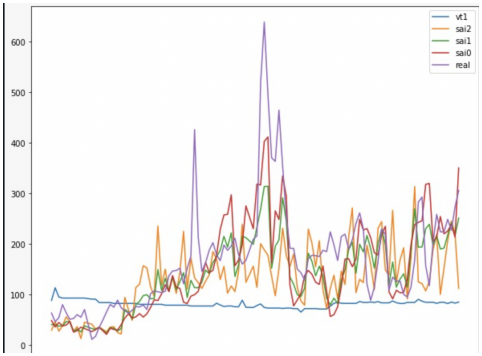
Model	R2	RMSE
Hierarchical Model (Baseline)	0.5011	52.78
Feed Forward (agg. features)	0.4573	55.05
Baseline + Feed Forward (agg. features)	0.5609	49.52
Baseline + Feed Forward (agg. features)	0.5758	48.67
Treatment 3	0.0009271	0.296

Table: Results

Hierarchical Baseline



Figure



Thank you