**Traffic Flow Prediction Dataset**

**Introduction**

The goal for this dataset is to forecast the spatio-temporal traffic volume based on the historical traffic volume and other features in neighboring locations. Specifically, the traffic volume is measured every 15 minutes at 36 sensor locations along two major highways in Northern Virginia/Washington D.C. capital region. The 47 features include: 1) the historical sequence of traffic volume sensed during the 10 most recent sample points (10 features), 2) week day (7 features), 3) hour of day (24 features), 4) road direction (4 features), 5) number of lanes (1 feature), and 6) name of the road (1 feature). The goal is to predict the traffic volume 15 minutes into the future for all sensor locations. With a given road network, we know the spatial connectivity between sensor locations.

**Processed Data**

Download link: [[Dataset](http://mason.gmu.edu/~lzhao9/pages/dataset_pages/datasets/spatial/traffic_dataset.mat)]

**Data format:** \*.mat (use Matlab to open)

**Data description:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable Name** | **Type** | **Size** | **Description** |
| tra\_X\_te | array of matrices | 1\*840 | test set input data: traffic indices for 840 contiguous quarter-hours  each element is a 36\*48 matrix: 36 spatial locations by 48 features |
| tra\_X\_tr | array of matrices | 1\*1261 | training set input data: traffic indices for 1261 contiouous quarter-hours  each element is a 36\*48 matrix: 36 spatial locations by 48 features |
| tra\_Y\_te | array of matrices | 36\*840 | * test set output data: traffic flowfor 36 locations in 840 contiguous quarter-hours from 2017-01-02 00:00 |
| tra\_Y\_tr | array of matrices | 36\*1261 | * training set output data: traffic flowfor 36 locations in 1261 contiouous quarter-hours until 2017-02-01 00:15 |
| tra\_adj\_mat | squared matrix | 36\*36 | adjacency matrix denoting the spatial connectivity of traffic network among 36 locations |
| **Variable Name** | **Type** | **Size** | **Description** |
| tra\_X\_te | array of matrices | 1\*840 | test set input data: traffic indices for 840 contiguous quarter-hours  each element is a 36\*48 matrix: 36 spatial locations by 48 features |

**Citation**

To use these datasets, please cite the papers:

Liang Zhao, Olga Gkountouna, and Dieter Pfoser. 2019. Spatial Auto-regressive Dependency Interpretable Learning Based on Spatial Topological Constraints. *ACM Trans. Spatial Algorithms Syst.* 5, 3, Article 19 (August 2019), 28 pages. DOI:https://doi.org/10.1145/3339823

**Contact**

Please contact Dr. Liang Zhao at [liang.zhao@emory.edu](mailto:liang.zhao@emory.edu).