CitiBike Demand Prediction by Mining Spatial-temporal Pattern

Yu Tang, Qian Xie

New York University

December 5, 2019

- Citi Bike is the largest bike-sharing system in New York City.
- 13,000 bikes and nearly 800 stations. Tens of thousands of rides every day.
- Concern: Supply shortage at major stations during rush hours.
- Bike demand prediction benefits bike rebalancing and management.
- Station-level demands have spatial and temporal dependencies.

Figure: Citi Bike

- Cluster-based prediction methods
- Spatial regression models
- Deep learning approaches
  - CNN
  - GCN
- We choose GCN for its power in recognizing patterns and dealing with graph-structured data.
- Unlike previous work, we consider not only historical data but also weather data and day of the week.

- Forecasting station-level demands in the next hour for each of the 128 bike stations in Lower Manhattan.
- Training data includes historical records, weather data and time factors.

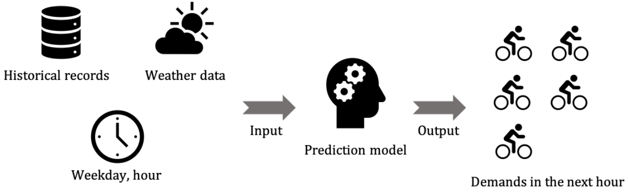


Figure: Prediction Problem

- Goal: Learn a function of features on a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$
- Input: $N \times T$ feature matrix X (N: number of nodes, T: number of input features) and an adjacency matrix A representing the graph structure
- Output: $N \times F$ feature matrix Z (F: number of output features per node)
- Every neural network layer can then be written as a non-linear function $H^{(l+1)} = f(H^l, A)$, with $H^0 = X$ and $H^L = Z$ (L: number of layers).

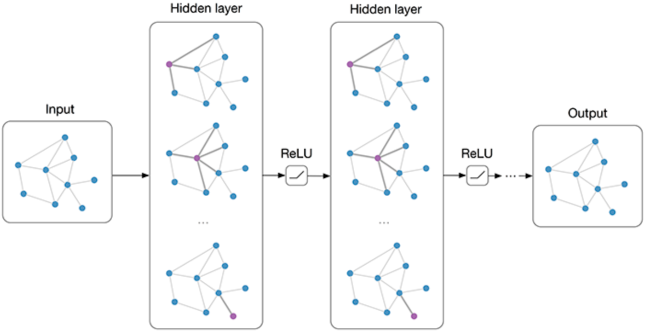


Figure: Multi-layer Graph Convolutional Network (GCN)

Methodology (Cont'd)

- Represent CitiBike data as graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ ($\mathcal{N}$: bike stations, $\mathcal{E}$: connectivity)
- Input features include historical hourly data (N: number of stations, T: hours).
- Adjacency matrix A: station dependency representation in terms of geographic distance, daily demand, average trip duration and demand correlation.
- GCN layer-wise propagation rule: $H^{(l+1)} = \sigma(D^{-\frac{1}{2}}AD^{\frac{1}{2}}H^{(l)}W^{(l)})$ ($\sigma(\cdot)$: non-linear activation function like the ReLU, D: diagonal matrix of A, $W^{(l)}$: matrix for the l-th neural network layer).
- Output of GCN + weather + day of the week = input of FCN.
- FCN layer-wise propagation rule: $H^{(l+1)} = \sigma(H^{(l)}W^{(l)})$.

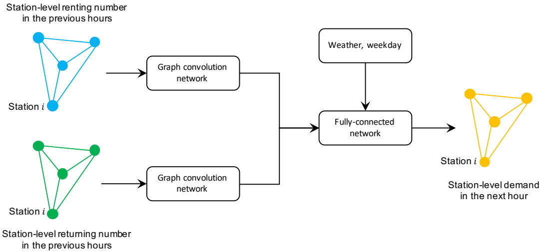


Figure: Methodology Framework

- Raw data is trip-based, including start time, end time, start station, end station, bike id and so on.
- Aggregate trip data from 2019/01/01 to 2019/10/31 to obtain station-level hourly rental/return data, daily OD demand and average trip duration.

Figure: Bike stations in Lower Manhattan

Table: Station-level hourly renting data

	Station ID			
	79	82	···	3812
2019/01/01 00:00	1	0	···	0
2019/01/01 01:00	0	0	···	0
⋮	⋮	⋮	⋮	⋮
2019/10/31 23:00	0	1	···	10

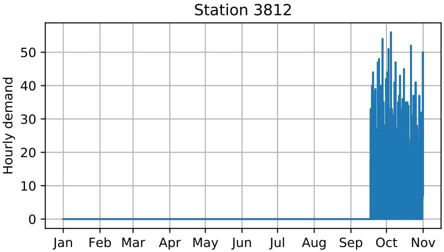
- Historical weather data are stored in the format of xml.
- Extract 1) temperature, 2) precipitation, 3) humidity, 4) heat index temperature, 5) wind speed and 6) weather condition description from the raw data.

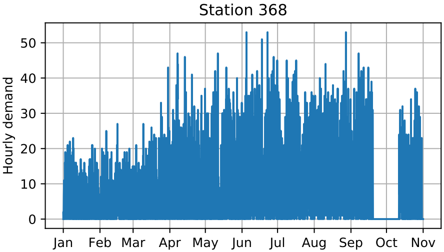
Table: Hourly historical weather data

	temp	precip	humidity	heat index	wind speed	weather condition
2019/01/01 00:00	0	9.3	89	2	10	heavy rain
2019/01/01 01:00	3	4.6	91	5	10	heavy rain
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2019/10/31 23:00	14	0.8	77	16	26	moderate rain

Data Preprocess

- Some stations are newly opened; some stations are temporarily "closed"





- Delete stations with more than 30% missing data
- Create a new binary feature "isMissing"

Impacts of weather

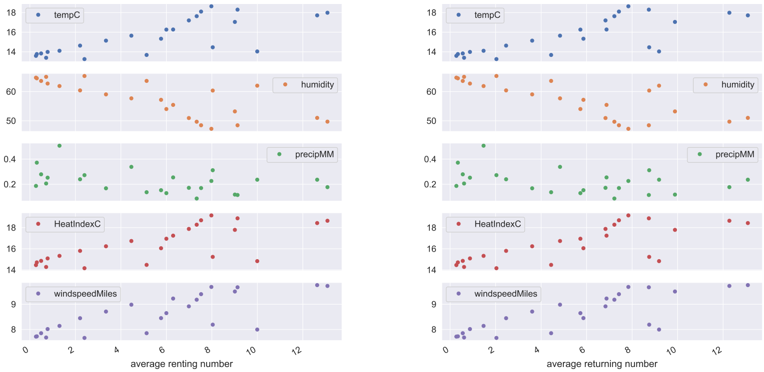


Figure: Relationship between average rental/return number and weather statistics

Temporal feature

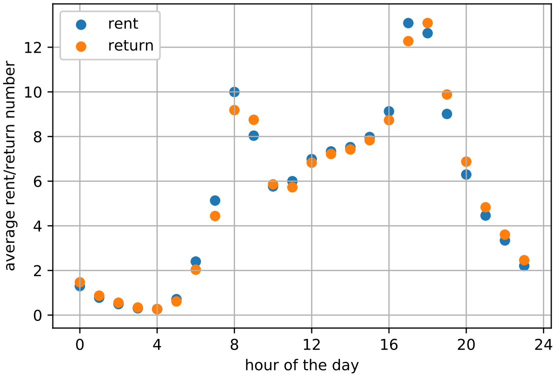


Figure: Average rental number and return number by hour of the day

Spatial dependency

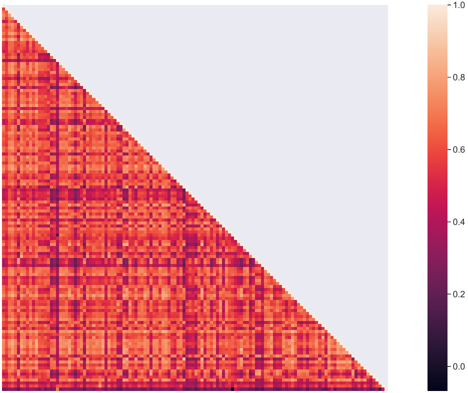


Figure: Correlation between station-level rental number

Four kinds of adjacency matrices

- Geometric distance (GD)
- Daily demand (DD)
- Trip duration (TD)
- Demand correlation (DC)

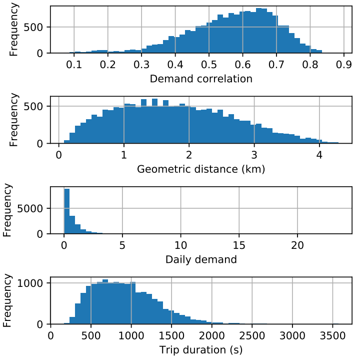


Table: Mean squared error

	GD	DD	TD	DC
Train	13.0	12.3	13.1	11.3
Test	18.4	18.0	18.3	17.9

Impacts of spatial-temporal features

- Threshold θ for adjacency matrix

Table: Mean squared error

θ	0.1	0.5	0.8
Train	13.9	13.3	11.3
Test	18.3	18.2	17.9

- Demand in T previous hours

Table: Mean squared error

T	1	2	3
Train	12.6	11.3	12.7
Test	18.9	17.9	18.2

Benchmarks

- Linear regression (LR)
- Multilayer perceptron (MLP)

Table: Mean squared error

	LR	MLP	GCN
Train	11.1	12.0	11.3
Test	22.6	19.3	17.9

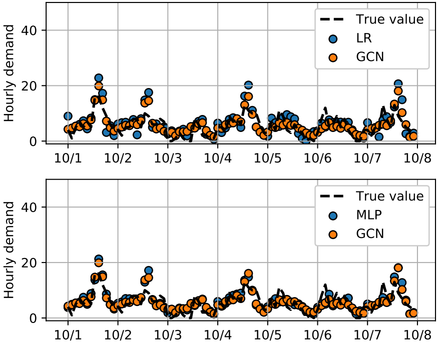


Figure: Prediction at station 79

- GCN application to the prediction of NYC Citi Bike spatial-temporal demand.
- Better performance compared to the benchmark (linear regression and multi-layer perceptron) in terms of Mean Squared Error (MSE) on
- Demand correlation matrix outperforms other three kinds of adjacency matrices.
- We take weather factor into account, it may be useful to consider social events and constructions.