Modeling and Understanding of Dynamic Human Mobility Patterns

Abhinav Jauhri

Thesis Committee:

John Paul Shen (Chair)

Anupam Datta

Jason Hong

Zhen Qian (CMU-CEE)

Electrical & Computer ENGINEERING

Manifestation Statement

Motivation: Human mobility is hard to understand.

Manifestation Statement

Motivation: Human mobility is hard to understand.

Ride sharing has become primary and economical mode of transportation.

- Didi completed 7 billion requests in China in 2017 [Ref: China Daily].
- In South Carolina 1% ride sharing fee has yielded more than a million dollars for municipalities to spend [Ref: NYTimes].

Manifestation Statement

Motivation: Human mobility is hard to understand.

Ride sharing has become primary and economical mode of transportation.

- Didi completed 7 billion requests in China in 2017 [Ref: China Daily].
- In South Carolina 1% ride sharing fee has yielded more than a million dollars for municipalities to spend [Ref: NYTimes].

> Exploit large scale human mobility data analytics to facilitate valuable services for societal good.

Contributions

- **Discovery of human mobility patterns** by leveraging extensive real world data.
  - Spatial and Temporal characterization

- **Modeling** for large scale synthetic data generation for broader research community.

- Learning **dynamic** mobility patterns in **real-time**.
  - Vehicle Placement Problem
  - Dynamic Pooling

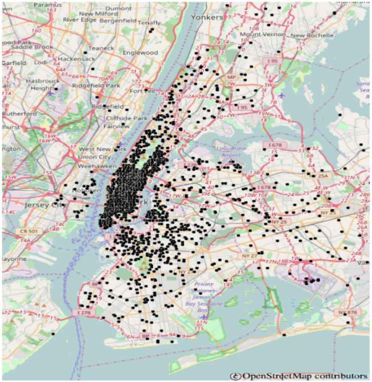
Electrical & Computer ENGINEERING

Contributions

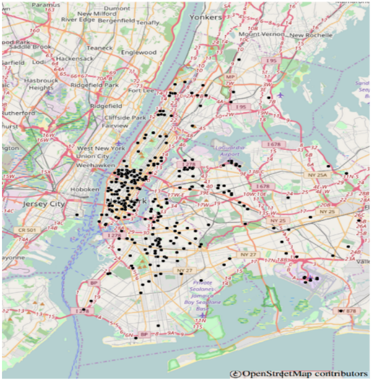
- **Discovery of human mobility patterns** by leveraging extensive real world data.
  - Spatial and Temporal characterization
- **Modeling** for large scale synthetic data generation for broader research community.
- Learning **dynamic** mobility patterns in **real-time**.
  - Dynamic Pooling
  - Vehicle Placement Problem

Electrical & Computer ENGINEERING

Human Mobility Patterns



(a) Between 8 - 8:05pm



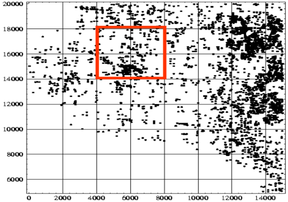
(b) Between 3 - 3:05am

Ride Requests in New York.

<u>Goal:</u> Characterize city level spatial and temporal variation

Electrical & Computer ENGINEERING

Spatial Variation – Fractal Dimension



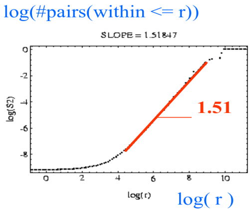


Figure: Self-similarity for cross roads of Montgomery county [Belussi 98]. $S_2 = \sum(\#\text{points})^2$.

Spatial Variation – Fractal Dimension

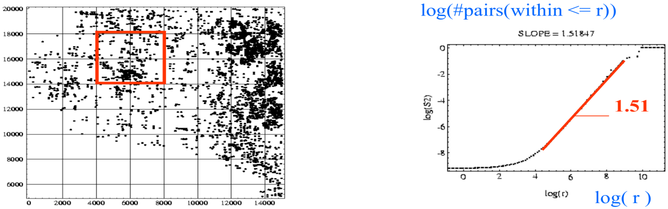
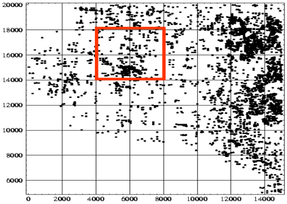


Figure: Self-similarity for cross roads of Montgomery county [Belussi 98]. $S_2 = \sum(\#\text{points})^2$.

Given a set of points $\mathbb{P}$ with finite cardinality and fractal dimension D_2, the average number of points within a square of radius ϵ follow a power law:

$$\overline{nb}(\epsilon) \propto \epsilon^{D_2} \qquad (1)$$

Spatial Variation – Fractal Dimension



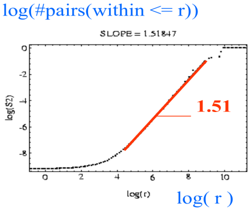
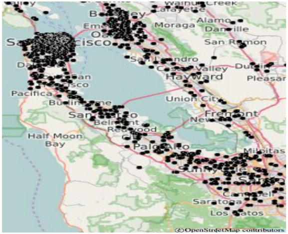


Figure: Self-similarity for cross roads of Montgomery county [Belussi 98]. $S_2 = \sum(\#\text{points})^2$.

Relevance of Fractal Dimension:

- Provides a way to characterize deviation from uniformity.
- Hypothesis testing and rule discovery.
- Other applications – query optimization for spatial access methods (SAM) [Faloustos 94].

Spatial Variation – Fractal Dimension



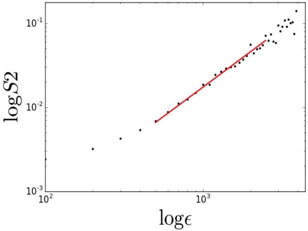


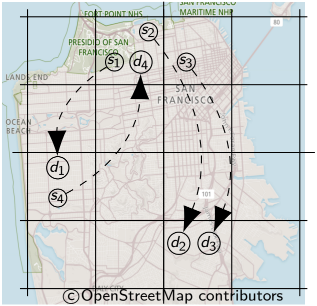
Figure: $S_2 = \sum(\#\text{points})^2$; Self-similarity for ride requests [Jauhri 17].

City	D_2 min.	D_2 max.	D_2 mean	fractal range (m.)
Chicago	1.003	1.459	1.206	(600, 3000)
Los Angeles	0.828	1.482	1.074	(1500, 4000)
New York	1.250	1.668	1.457	(450, 2500)
San Francisco	1.049	1.686	1.343	(450, 2500)

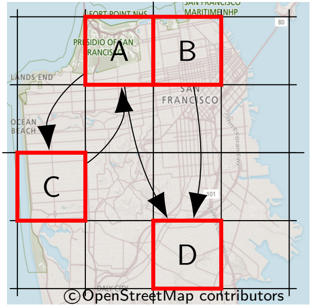
Table: Summary of measured correlation fractal dimension (D_2) for four cities; computed over a week for every 3-minute time snapshot.

Discovery: Number of ride requests within a bounded of region follow a power law.

Temporal Variation – Ride Request Graph



(a) Four ride requests distributed spatially over a map

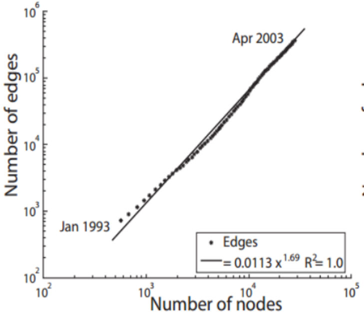


(b) Corresponding Ride Request Graph with four nodes (marked by red boxes) and directed edges.

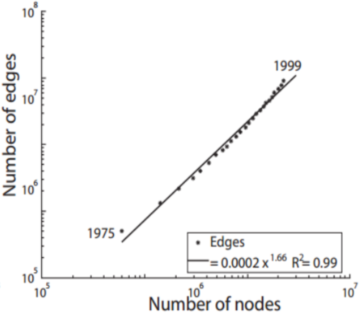
Figure: Transformation of ride requests for a small interval into a directed **Ride Request Graph (RRG).**

Temporal Variation – Densification Power Law

Time-evolving graph like arXiv citation graph, the Patent citation graph, social network graph, and many others share the **Densification Power Law property** [Chakrabarti 98].



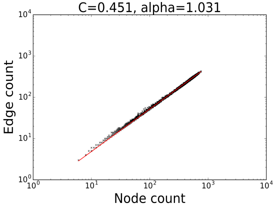
(a) arXiv



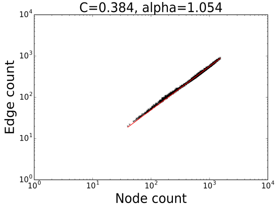
(b) Patents

Image source: Chakrabarti, D. (2006). Graph mining: Laws, generators, and algorithms.

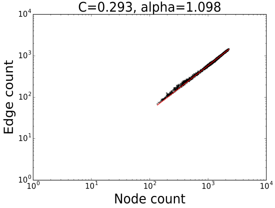
DPL Property – (Number of ride requests) $\propto$ (Number of grids)$^{\alpha}$ [Jauhri 17].



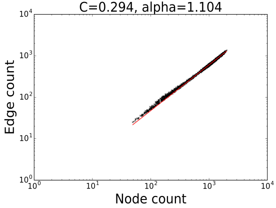
(a) Hyderabad



(b) Paris



(c) New York



(d) San Francisco

Densification Power Law & Temporal Variation

Discovery: Ride Request Graphs obey Densification Power Law property over time.

DPL and Poolability

Poolability – Percentage of ride requests which:

- originate within the same time snapshot of $5 minutes$.
- pickup location within circle of radius $\epsilon_{sr} = 100m$.
- drop-off location within a circle of radius of $\epsilon_{dr} = 1000m$.

DPL and Poolability

Poolability – Percentage of ride requests which:

- originate within the same time snapshot of $5 minutes$.
- pickup location within circle of radius $\epsilon_{sr} = 100m$.
- drop-off location within a circle of radius of $\epsilon_{dr} = 1000m$.

City	*Mean Poolability*	α
Hyderabad	2.23	1.031
Paris	2.39	1.054
New York	4.48	1.098
San Francisco	5.48	1.104

Mean poolability, and α for four cities computed for over a week's data.

Electrical & Computer ENGINEERING

Summary of Mobility Patterns

1. Fractal dimensionality provides an approximation of how ride requests are geographically distributed.

Summary of Mobility Patterns

1. Fractal dimensionality provides an approximation of how ride requests are geographically distributed.

2. RRGs provide a rigorous model to characterize requests over time.

Summary of Mobility Patterns

1. Fractal dimensionality provides an approximation of how ride requests are geographically distributed.

2. RRGs provide a rigorous model to characterize requests over time.

These characterizations could help in synthetic generation and to understand how algorithms perform for applications related to human mobility.

Contributions

- **Discovery of human mobility patterns** by leveraging extensive real world data.
  - Spatial and Temporal characterization
- Modeling for large scale synthetic data generation for broader research community.
- Learning **dynamic** mobility patterns in **real-time**.
  - Vehicle Placement Problem
  - Dynamic Pooling

Electrical & Computer ENGINEERING

Synthetic Data Generation – First Attempt

Used PoI data from OSM to act as proxy for spatial distribution of population, and a simple model to construct graph which obeys DPL property.

Synthetic Data Generation – First Attempt

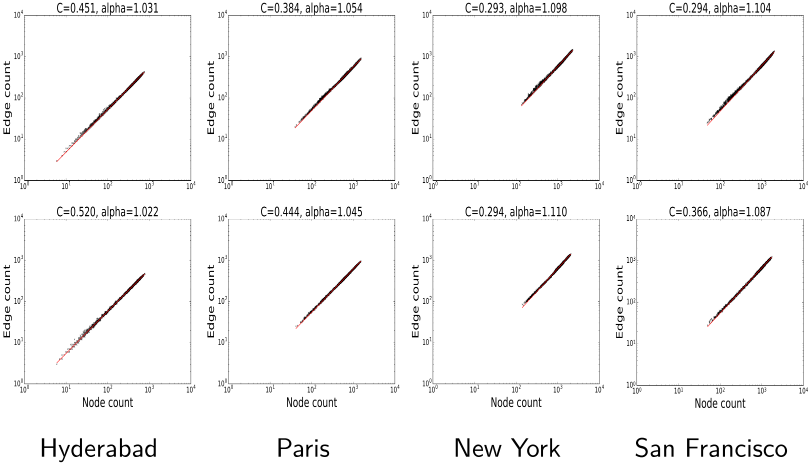


Figure: DPL plots from real data (top row) and synthetic data (bottom row) for four cities.

Synthetic Data Generation – Poolability Comparison

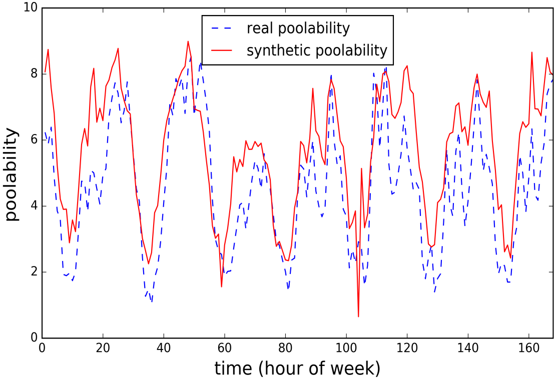


Figure: Comparison of poolability generated by synthetic data (red line) and real data (dotted blue line).

Synthetic Data Generation – GANs (Initial Results)

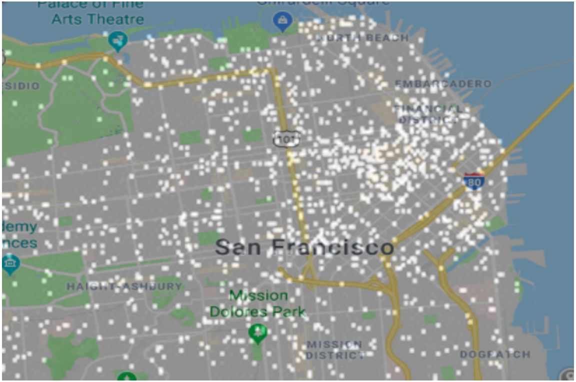


Figure: Ride requests generated for a 5 minute time interval at 11am

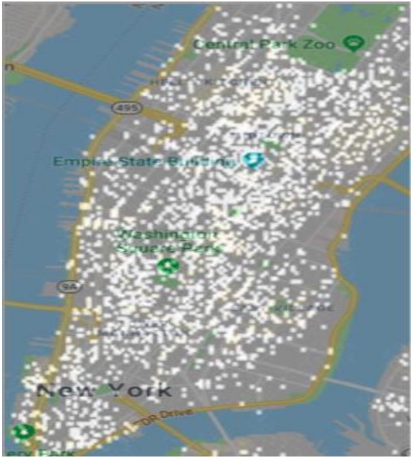


Figure: Ride requests generated for a 5 minute time interval at 11pm

Contributions

- **Discovery of human mobility patterns** by leveraging extensive real world data.
  - Spatial and Temporal characterization
- **Modeling** for large scale synthetic data generation for broader research community.
- Learning **dynamic** mobility patterns in **real-time**.
  - Vehicle Placement Problem
  - Dynamic Pooling

Electrical & Computer ENGINEERING

Learning in Dynamic Environments

- Real-time mobility environment is changing; drop-off and pickup locations are hard to predict.
- Costs alter including traffic conditions, rider and driver demands.

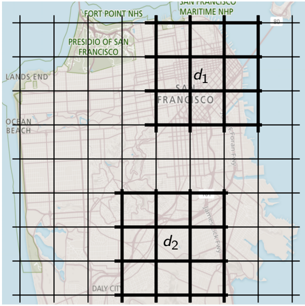
<u>Goal:</u> Learn to make real-time decisions in a dynamic environment at city-scale.

Vehicle Placement Problem

How to reduce rider waiting and driver idling time by placing vehicles close to riders without knowledge of future ride requests?

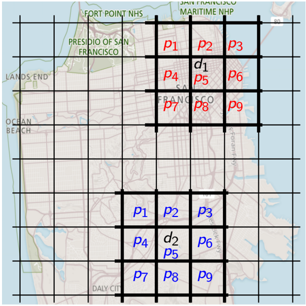
In San Franciso, it's estimated that approximately 20 percent of the miles traveled by Uber and Lyft drivers are without passengers. – Citylab, April, 2018.

Vehicle Placement Problem



- d_i - dropoffs at time snapshot t

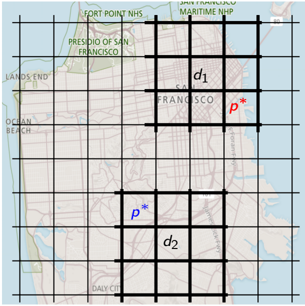
Vehicle Placement Problem



- d_i - dropoffs at time snapshot t
- p_i - possible placements for d_1 by time snapshot $t+1$
- p_i - possible placements for d_2 by time snapshot $t+1$

Electrical & Computer ENGINEERING

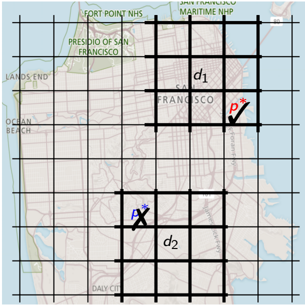
Vehicle Placement Problem



- d_i - dropoffs at time snapshot t
- p^* - placement for d_1 by time snapshot $t+1$
- p^* - placement for d_2 by time snapshot $t+1$

Two placements are made using some algorithm.

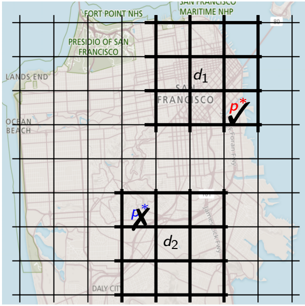
Vehicle Placement Problem



- d_i - dropoffs at time snapshot t
- p^* - placement for d_1 by time snapshot $t+1$
- p^* - placement for d_2 by time snapshot $t+1$

Two placements are made using some algorithm.

Vehicle Placement Problem



- d_i - dropoffs at time snapshot t
- p^* - placement for d_1 by time snapshot $t+1$
- p^* - placement for d_2 by time snapshot $t+1$

Two placements are made using some algorithm.

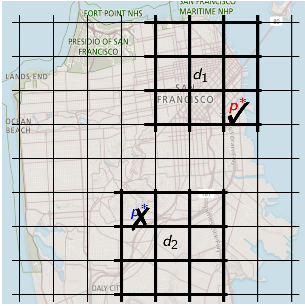
Reward R is computed for every time snapshot:

$$R(t+1) = \frac{\#\text{good placements}}{\#\text{total placements}}$$

For the example above:

$$R(t+1) = \frac{1}{2}$$

Vehicle Placement Problem



- d_i - dropoffs at time snapshot t
- p^* - placement for d_1 by time snapshot $t+1$
- p^* - placement for d_2 by time snapshot $t+1$

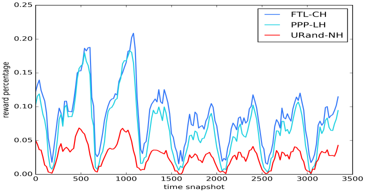
Two placements are made using some algorithm.

Reward R is computed for every time snapshot:

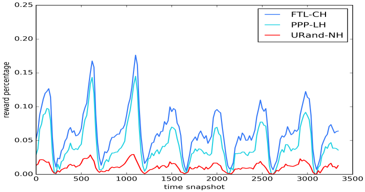
$$R(t+1) = \frac{\#\text{good placements}}{\#\text{total placements}}$$

Objective: Maximize the reward R over time.

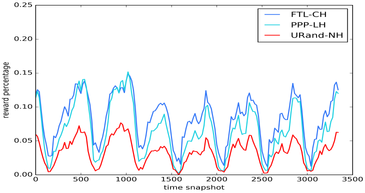
Vehicle Placement Problem – Results



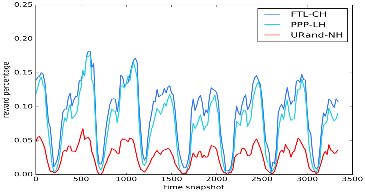
(a) Chicago



(b) Los Angeles



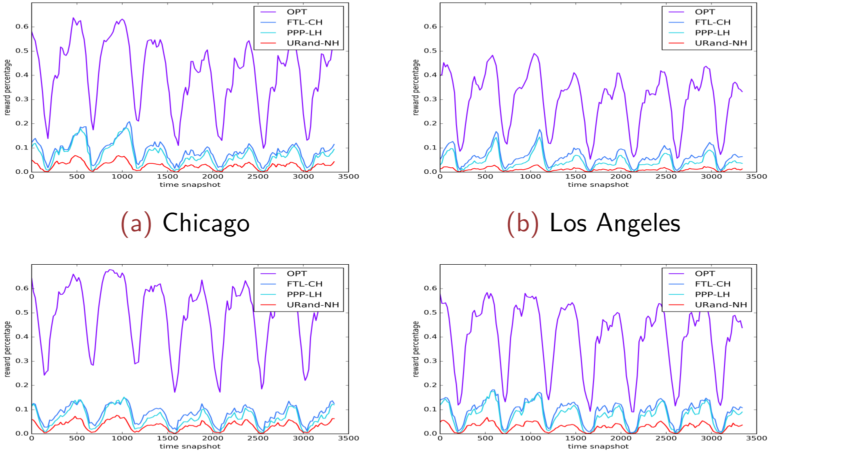
(c) New York



(d) San Francisco

Figure: Reward percentage plots for a week with three minute time snapshots [Jauhri 17].

Vehicle Placement Problem – Results



(a) Chicago

(b) Los Angeles

(c) New York

(d) San Francisco

Figure: Reward percentage plots for a week in comparison with optimal.

Electrical & Computer ENGINEERING

Ideas to Improve Performance of the Vehicle Placement Problem

- Perform placements for beyond $t + 1$ by using reinforcement learning.

Ideas to Improve Performance of the Vehicle Placement Problem

- Perform placements for beyond $t + 1$ by using reinforcement learning.
- Reinforcement Learning methods are difficult to train in practice; how can we learn model(s) deployable at city scale?

Electrical & Computer ENGINEERING

Ideas to Improve Performance of the Vehicle Placement Problem

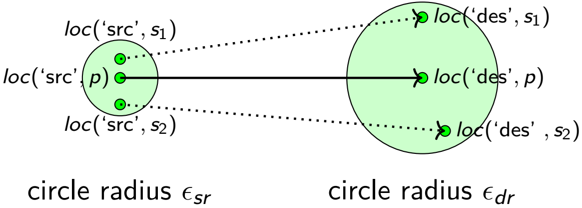
- Perform placements for beyond $t + 1$ by using reinforcement learning.
- Reinforcement Learning methods are difficult to train in practice; how can we learn model(s) deployable at city scale?
- How useful is historical information or is real-time information enough?

Electrical & Computer ENGINEERING

Dynamic Pooling

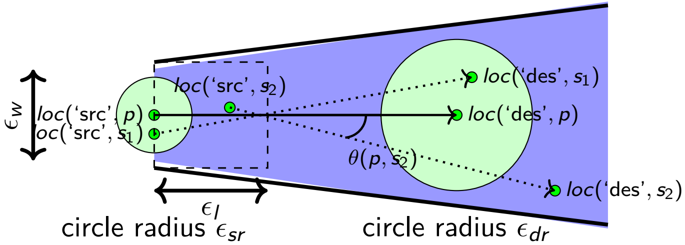
What is the design space of to perform dynamic pooling to increase average vehicle occupancy?

Dynamic Pooling – Restricted Constraints



Restricted Constraints: The source region is defined by a circle with radius ϵ_{sr} and centers at the pick up point of the primary request p. The destination region is defined by another circle with radius ϵ_{dr}.

Dynamic Pooling – Hybrid Constraints



Hybrid Constraints: An expanded set of constraints to do pooling [Jauhri 17].

Dynamic Pooling – Comparison of Techniques

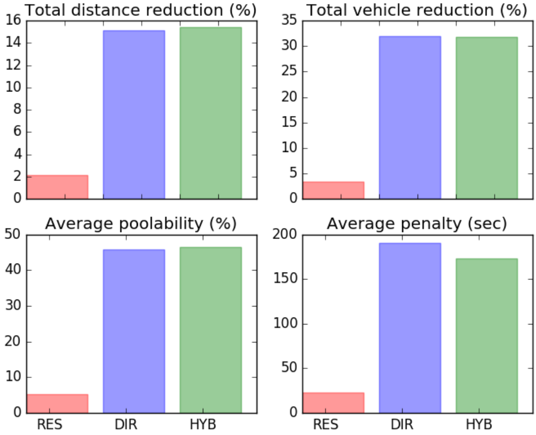


Figure: Comparison of pooling using Restricted, Directed, and Hybrid constraints in San Francisco.

Dynamic Pooling – Comparison of Cities

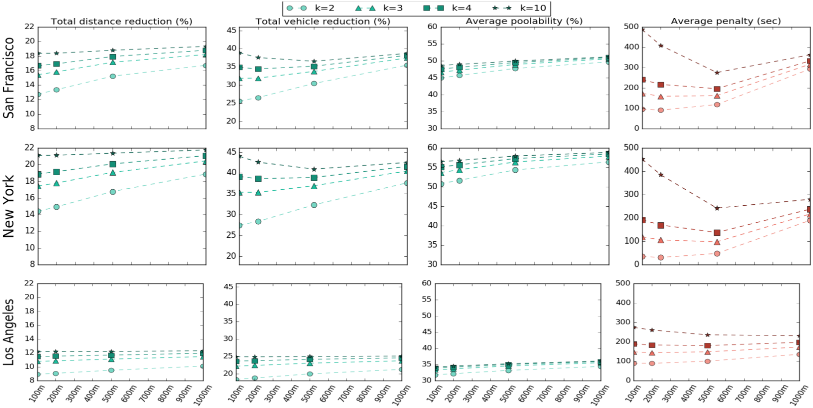


Figure: Results with varying e_{sr}, k for three benefits and one cost metrics. $\epsilon_{dr} = 1000m, \epsilon_w = 2000m, \epsilon_\theta = 20$ are kept constant.

Dynamic Pooling – Societal Benefits

Metric	*San Francisco*	*New York*	*Los Angeles*	*Mean*[1]
Total Travel Distance Reduction (%)	17.13	19.06	11.01	15.76
Total Vehicle Count Reduction (%)	33.76	36.93	23.03	31.23
Mean Poolability (%)	48.94	56.39	34.52	46.61
Mean Travel Time Penalty (sec)	162.12	97.55	148.17	135.94

Table: Summary of benefits and costs.

[1] Across 3 cities.

Things to be done

- Generate synthetic data for small time instances (≈ 5 minute) and city scale.
- Rigorously validate synthetic data using spatial and temporal properties.
- Develop techniques to attain close to optimal placement of vehicles.
- Find societal benefits of pooling and placement by modulating the volume of requests (what-if scenarios).
- Other ideas...

List of Publications

- Jauhri, Abhinav, Stocks, Brad, and Shen, John Paul. Generating Large Scale Mobility Patterns *(working paper)*.
- Jauhri, Abhinav, Nuanes, Tyler and Shen, John Paul. Reinforcement Learning for Vehicle Placement *(working paper)*.
- Jauhri, Abhinav, et al. From Millions of Check-ins To Candidate Pins: Finding the Needle in a Haystack *(working paper)*.
- Chen, Xinlei, Jauhri, Abhinav, et al. Actuation System for City-wide Ride-based Vehicular Crowdsensing *(in submission)*.
- Jauhri, Abhinav, et al. "Space-Time Graph Modeling of Ride Requests Based on Real-World Data." arXiv preprint arXiv:1701.06635 (2017).
- Chen, Min Hao, Jauhri, Abhinav, and John Paul Shen. Data Driven Analysis of the Potentials of Dynamic Ride Pooling, 10th ACM SIGSPATIAL Workshop on Computational Transportation Science (IWCTS'17).
- Jauhri, Abhinav, Carlee Joe-Wong, and John Paul Shen. "On the Real-Time Vehicle Placement Problem." arXiv preprint arXiv:1712.01235 (2017).
- Jauhri, Abhinav, Bradley McDanel, and Chris Connor. Outlier detection for large scale manufacturing processes. Big Data, 2015 IEEE International Conference on. IEEE, 2015.
- Jauhri, Abhinav, Martin Griss, and Hakan Erdogmus. Small Polygon Compression For Integer Coordinates. arXiv preprint arXiv:1509.05505 (2015).
- Erdogmus, H., et al. Opportunities, Options, and Enhancements for the Wireless Emergency Alerting Service. Carnegie Mellon University, Technical Report CMU-SV-15-001 (2015).
- Jauhri, Abhinav, et al. A comparison of antenna placement algorithms. Annual Conference on Genetic and Evolutionary Computation. ACM, 2014.

Electrical & Computer ENGINEERING

Contributions

- **Discovery of human mobility patterns** by leveraging extensive real world data.
  - Spatial and Temporal characterization
- **Modeling** for large scale synthetic data generation for broader research community.
- Learning **dynamic** mobility patterns in **real-time**.
  - Vehicle Placement Problem
  - Dynamic Pooling