Geometry of Traffic

Faculty Mentor: Prof. Richard Sowers, Team Leaders: Daniel Carmody, Kavjit Durairaj, Sarah Mousley-Mackay
Scholars: Dipro Ray, Fengdeng Lyu, Hyunsoo Lee, Jing Ren, Bessie An, Yu Zhang, Charles Liu, Jerry Li, Andong Jing

Introduction

Motivations

Major cities are increasingly repurposing curb space once reserved for parking meters to serve food trucks, bike sharing programs, and public transit vehicles. To reduce frustration caused by the decrease in supply of curbside parking, understanding the current parking behavior of individual drivers is a key input to making effective change.

Current Goals

- Find periodicity of the data in the time domain; cluster the data to find spatial patterns
- Understand relationship between parking usage and factors such as ride sharing services, schools, restaurants and their typical hours of operations.

Data Descriptive Statistics

- **Source:** publicly available from SFMTA (http://sfpark.org/about-the-project/pilot-evaluation/)
- **Data size:** Over 19 million lines, 7,000 of 28,800 metered spaces and 12,250 spaces in 15 of 20 parking garages
- **Time frame:** 04/01/2011 – 05/31/2013
- **What it has:** Occupancy of metered parking spots at given times.

Low Rank Factorization

Let A be an $n \times m$ matrix. A low rank factorization of A is a decomposition $A \approx WH$ where W is $n \times k$ and H is $k \times m$ with $k \ll \min(n, m)$. The precise sense in which $A \approx WH$ is measured by the cost function $\|A - WH\|_{fro}$, the norm of the difference when we just think of the matrices as vectors in $\mathbb{R}^{n \times m}$.

Because our data represents occupancy in parking lots, which is always a non-negative number, we require that our low-rank factorization WH consist of non-negative matrices. It was also shown in [1] that compared to methods like principal component analysis, non-negative matrix factorization does a better job of learning human-identifiable features.

Periodic Features

Periodic Projection:

To understand the most important daily parking features, we perform a rank 3 factorization on a periodization of the data and plot the features as occupancy vs time of day in 20-minute intervals. We discover that our factorization separates parking on a temporal basis into morning, mid-day, and evening parking.

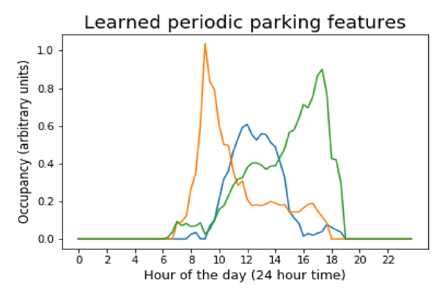

Figure 1: Learned Periodic Parking Features

Clustering Parking Data

Motivation and Implementation
Looking at the parking data of San Francisco, we hypothesized that, apart from periodicity, the influences of other factors might cause similar parking patterns in groups of parking spots. How do we find these factors?

To answer this question, we clustered the columns of H matrix (of the non-negative matrix factorization) using a standard clustering algorithm, and investigated the clusters that form. We implemented spectral clustering with different distance functions (Euclidean Distance, Cosine Similarity) and tuned gamma parameters.

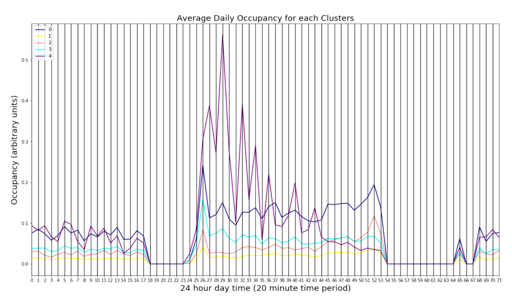

Figure 2: Clusters in Occupancy vs Time graph

Analysis of Results
Using ratiocut method to find the optimal cluster size, we performed spectral clustering to separate all San Francisco parking lots into five clusters. To visually inspect the clusters, we plotted all parking spots color-coded by their cluster on a map. To look into cluster prevalence per district, we looked at the distribution of clusters among San Francisco districts using pie charts. Surprisingly, the largest cluster (by magnitude) occurred away from popular districts - explained by the fact that our dataset only cover street meters and not garages and indoor parking lots.

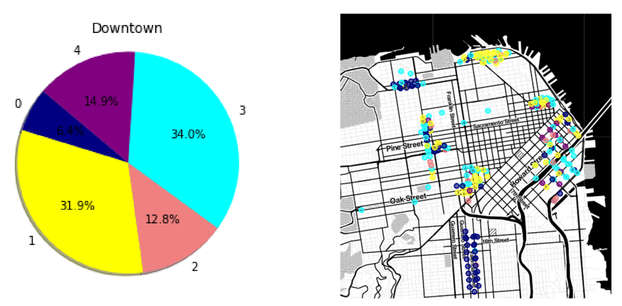

Figure 3: *Parking Spots Clustering*

Dynamic Pricing Visualization

Overall Visualization:
We created visualizations using folium circle map, which show the changes of occupancy rate under first price adjustment. Even though pricing went through minor changes, the occupancy did spread more evenly, indicating improved parking availability.

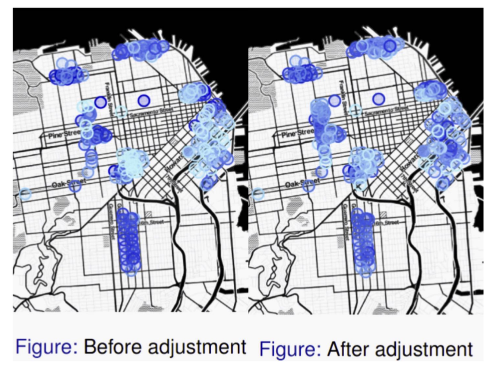

Figure 4: Occupancy Rate Change After Price Adjustment

Specific Region and Weekday To further investigate the effectiveness of demand-responsive pricing, we checked on specific region and weekday. The first graph we generate is an x-y plot on how long it takes for a certain parking space to reach 60% (Green), 70% (Yellow) and 80% (Red) of the maximum occupancy.

This is an example for Monday, Downtown area. We can clearly see that, there are more dots above the x = y line, which means that, on average, it takes longer time for a parking space to reach certain occupancy after policy change, indicating the policy successfully improved parking availability.

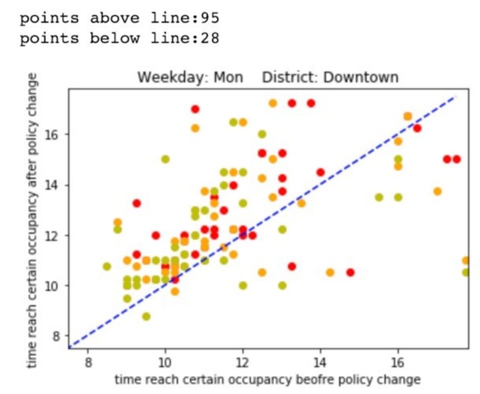

Figure 5: Time vs. 60%, 70%, 80% Occupancy

Then, we linked the three points of a parking space together, so we could visualize a continuous deformation of improvements. According to these graphs, more than half of the parking spaces are under positive influence of the pricing policy, while some of the blocks are not quite as responsive.

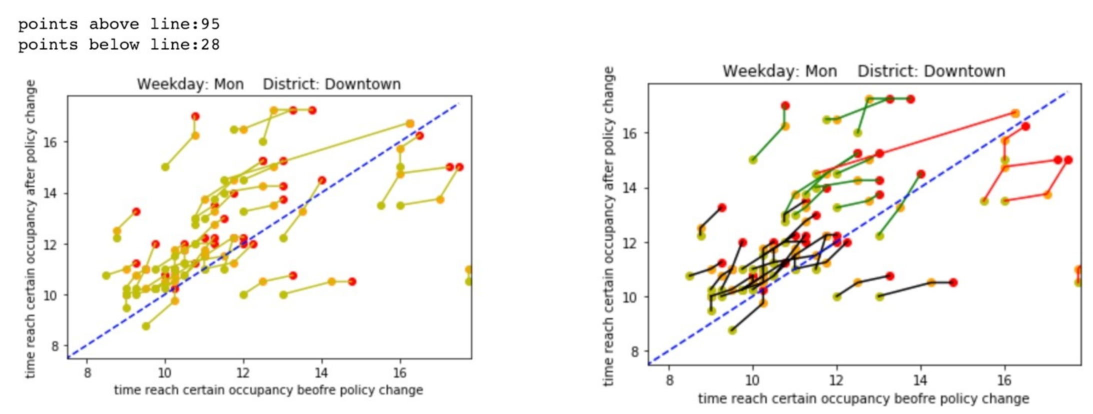

Figure 6: *K-Means Clustering*

We took the coordinates of these 3 points as features and applied K-Means algorithm to generate 3 clusters in our graphs. The policy is proved effective for the "green" parking lots but perform poorly on the "red" ones. In the following step, we are going to dig deeper into the location and economic data around the parking lot to see why these parking spaces are clustered together, and how we can change the policy to further improve parking availability.

Regression and Occupancy Visualization

Nearby Stores
We retrieved the number of stores, restaurants, movie theaters, bars, etc. within 200 meters of a parking lot using the Google API. A linear regression model of these features was fitted against the the percentage usage (total cars parked in a day / capacity) in log scale. This parking data is derived from October 2011 every Monday to Saturday.

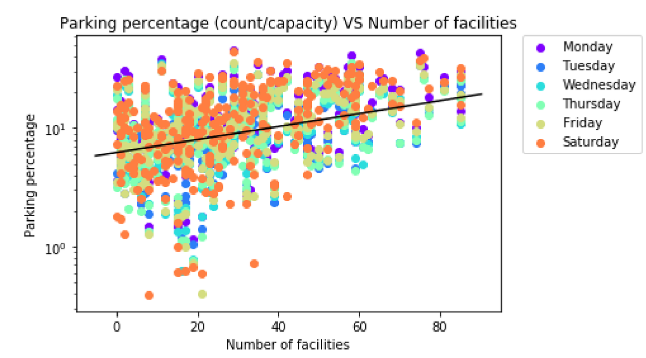

Figure 7: Parking Percentage VS Number of Facilities Nearby

Impact of Ride-sharing Services
We tried to understand how the structure of the nearby environment might affect parking usage in a given area. For example, we studied how ride sharing service usage and number of restaurants nearby could affect the parking behavior. To help look for patterns, we are building an interactive visualization to display the occupancy of parking lots for a given day of the week and hour.

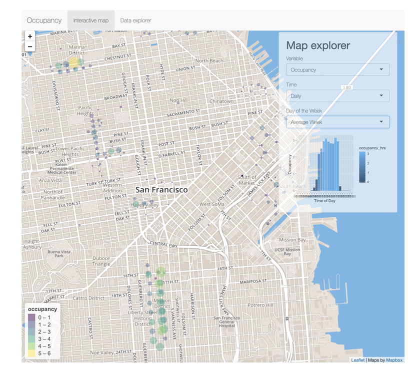

Figure 8: The interactive Visualization of the Occupancy of Parking Lots

Future Directions

- Expand the initial data set and test different hyper-parameters in our factorization and clustering methods
- Building on preliminary results using Kaplan-Meier estimators, dynamic pricing team will attempt to fit a series of Cox Models with more complex covariates under a mixed censoring context
- Identify areas with the most and least geographic coverage of parking services and provide recommendations based on popular availability of ride sharing services in those areas

References

[1] Lee, Daniel and Seung, H. Sebastian. Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 788â791 (21 October 1999).

Support for this project was provided by the Illinois Geometry Lab and the Department of Mathematics at the University of Illinois at Urbana-Champaign.