# Locational Demand

Tanner Fiez

1   Spatial and Temporal Heterogeneity

Parking occupancy across the Belltown neighborhood in Seattle displays both temporal and spatial heterogeneity. Figure 1a demonstrates the spatial heterogeneity by looking at the average occupancy load for each block in the neighborhood, and figure 2a demonstrates the temporal heterogeneity by looking at the average occupancy over the entire neighborhood at each time of day and day of week that paid parking is available.
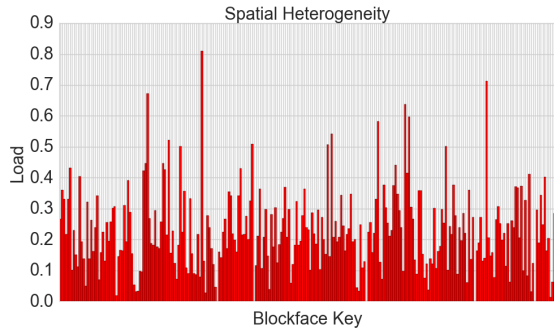
(a) Spatial heterogeneity plot of the Belltown neighborhood.
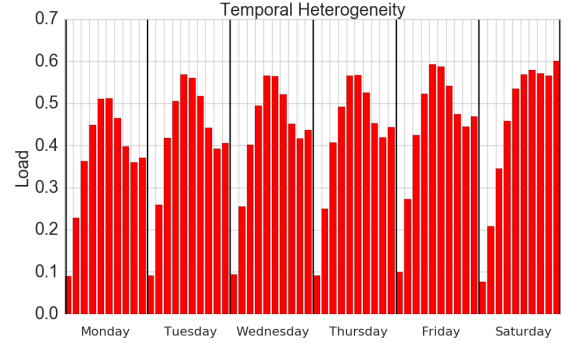
(b) Temporal heterogeneity plot of the Belltown neighborhood.
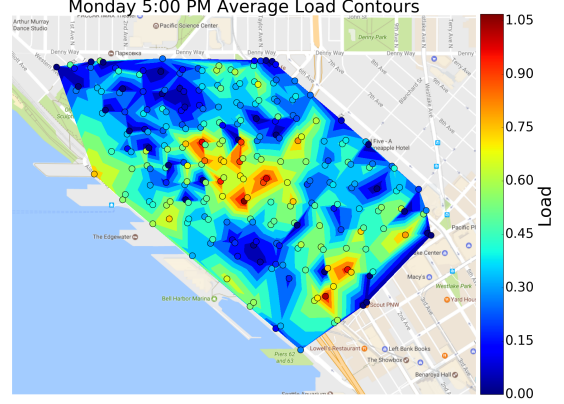
Figure 1

The spatial heterogeneity demonstrated in figure 1a, is one of the fundamental reasons that using average occupancy as the sole factor in policy decisions is an insufficient metric when considering how occupancy impacts congestion. The temporal heterogeneity demonstrated in figure 1b, shows that parking behavior is similar during weekdays, while behavior on Saturdays is very different. These properties of parking behavior are key when designing control strategies to mitigate congestion. We further observe that within the Belltown neighborhood there are centers of demand that dictate a majority of the congestion. The centers of demand are repeatable at given time of day and day of week over time, and we also find that the centers of demand change with time of day and day of week. We model locational demand by using a Gaussian Mixture Model (GMM) with spatial and occupancy data.

2   Gaussian Mixture Model for Locational Demand Analysis

As discussed previously, the occupancy of parking exhibits spatial heterogeneity within the Belltown neighborhood. By using a GMM we can identify centers of demand based on spatial information and occupancy data. By first considering only the occupancy data we can begin to get an idea for the distribution of parking occupancy and how parking behavior changes over the course of a day.
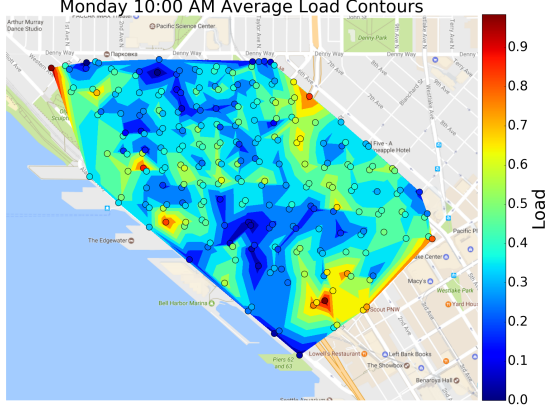
(a) 10 Contours of the average load in the Belltown neighborhood at 10:00 AM on Monday.

(b) 10 Contours of the average load in the Belltown neighborhood at 5:00 PM on Monday.

Figure 2

Although in figure 1a it can be observed that across Belltown there is a similar average load on Monday at 10:00 AM and 5:00 PM, the contour plot reveals that locational demand changes over a course of a day. In the morning parking is spread more uniformly across the neighborhood as people arrive to work. In the evening the parking becomes much more concentrated in the center of the neighborhood where many of the popular restaurants and bars are concentrated.

Using a GMM we can model the centers of demand and give an estimate the influence the centers of demand have on neighboring blockfaces. A GMM can be used as an unsupervised clustering method to find partitions in a population containing similar samples. We consider a Multivariate GMM in which each sample $\bar{x}_i$ is a vector containing spatial and occupancy data as $\bar{x}_i = [x_{\text{latitude}},\ x_{\text{longitude}},\ x_{\text{load}}]$. The purpose is to find which of the K possible mixture components each sample belongs to with the highest probability. Each mixture component has a prior on the probability of a sample belonging to it, $P(C_j) = \pi_j$. The likelihood of a sample belonging to a mixture component is then given as a Multivariate Gaussian Distribution

$$P(\bar{x}_i|C_j) = \frac{1}{(2\pi)^{\frac{3}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\bar{x}_i-\bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x}_i-\bar{\mu}_j)}. \tag{1}$$

We assume that the features ($x_{\text{latitude}},\ x_{\text{longitude}},\ x_{\text{load}}$) are conditionally independent given the component, i.e. each covariance matrix Σ_j is diagonal. The GMM is a classic chicken and the egg problem. We want to find the component that each sample belongs to, but to do so we need to know the distributions of each of the components. This problem is solved by the expectation-maximization (EM) algorithm which maximizes the log likelihood of the data. In the EM algorithm we initialize each of the component parameters, i.e. the prior, mean, and covariance, randomly. Until convergence is reached, which occurs when the complete log likelihood given by $L = \sum_{i=1}^{n} log \sum_{j=1}^{k} P(C_j)P(\bar{x}_i|C_j)$, increases by an amount smaller than some ϵ, two steps are repeated at each iteration of the algorithm. These are the E step, in which the expected values of the unobserved component labels are calculated with the current parameter values, and the M step, in which new parameter values are calculated to maximize the probability of the data. The E step for all samples x_i is

$$P(C_j|\bar{x}_i) = \frac{P(C_j)P(\bar{x}_i|C_j)}{\sum_{i'} P(C_{i'})P(\bar{x}_i|C_{i'})} \tag{2}$$

and the M step for all components is

$$P(C_j) = \frac{1}{n} \sum_{i=1}^{n} P(C_j|\bar{x}_i) \tag{3}$$

$$\bar{\mu}_i = \frac{\sum_{i=1}^{n} \bar{x}_i P(C_j|\bar{x}_i)}{\sum_{i=1}^{n} P(C_j|\bar{x}_i)} \tag{4}$$

2

$$\Sigma_j = \frac{\sum_{i=1}^{n}(\bar{x}_i - \bar{\mu}_j)^2 P(C_j|\bar{x}_i)}{\sum_{i=1}^{n} P(C_j|\bar{x}_i)}. \tag{5}$$

At convergence each sample x_i is assigned the component label which maximizes the posterior probability $P(C_j|\bar{x}_i)$, and the component parameters are recorded.

The GMM also involves a model selection problem in selecting the appropriate number of mixture components. We used both the Akaike Information Criterion and the Bayesian Information Criterion when choosing the model. The Akaike Information Criterion attempts to find the model that describes the unknown reality most adequately, while the Bayesian Information Criterion
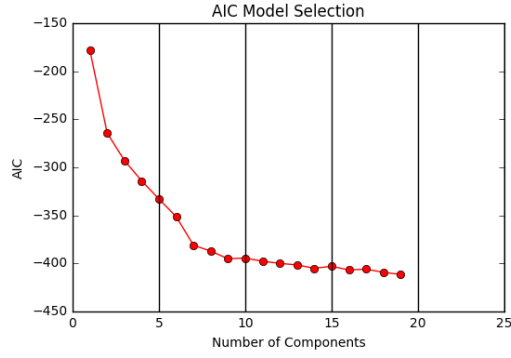
(a) Akaike information criterion for model selection.
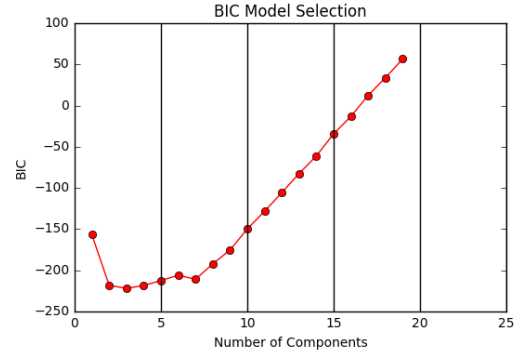
(b) Bayesian information criterion for model selection.

Figure 3