Assigning Awaiting Area for Food Delivers
- Final Report

Team Members

Yizhan Li

Tongyan Feng

2022 Fall Semester

EN.560.653

Instructor: Professor Lauren Gardner

1. Problem Description

1.1 Background

During dashers' (delivers work for Doordash) daily work, they pick up items from the restaurant and deliver the items to the customers. During their idle time, dashers may just return home or go somewhere to wait for the next order. However, their awaiting places can influence their response time to the coming orders and the number of coming orders, for the Doordash's algorithm will forward the orders to the nearest idle delivers.

1.2 Target of this project

In this project, the team is planning to recommend awaiting/rest points for the delivers, the points radiate areas to ensure the dashers can receive and react to new orders rapidly, these awaiting areas may have high level of closeness or betweenness centralities to provide higher response efficiency than other locations. With the help of assigned awaiting areas, the deliver can have a better place, where is generally closer to the next pickup locations and have higher possibility to receive orders, to stay during their spare time between orders.

1.3 Literature Review

In order to conduct the research with more theoretical support, also to learn from the former researches in this area, the team went through some relative scholars. The first essay is by Liu. In Liu's paper, they explore a framework of how cabs receive delivery orders during their free time and plan the optimal delivery paths with drop-off points, by using the construction algorithm and the Adaptive Large Neighborhood Search (ALNS) algorithm (Liu 1031). An essay by Li about clustering in this kind of problem also helped the team a lot. Li states that it can be effective to obtain community structures by the machine learning algorithms, such as using K-means to cluster the low-dimensional feature matrix (Li 1). They support us with the basic concept and algorithms to build some nodes can help deliveryman handle with multiple orders generated within a small range of time intervals.

2. Data Acquisition

The original dataset contains location information (coordinates, address) of over 10,000 restaurants in New York, Boston and Washington D.C. and the restaurants' review counts. The original size of the dataset is 11219x16, to scale down the problem, the team concentrate on a region in Brooklyn, New York, where the ZIP code starts with 11221. The downsized dataset's size became 124x6, only contains the information of 124 restaurant in the focus region.

3. Network Formulation

As this problem is generally a geographic problem, the team used ArcGIS to visualize the network (shown in Figure 1).

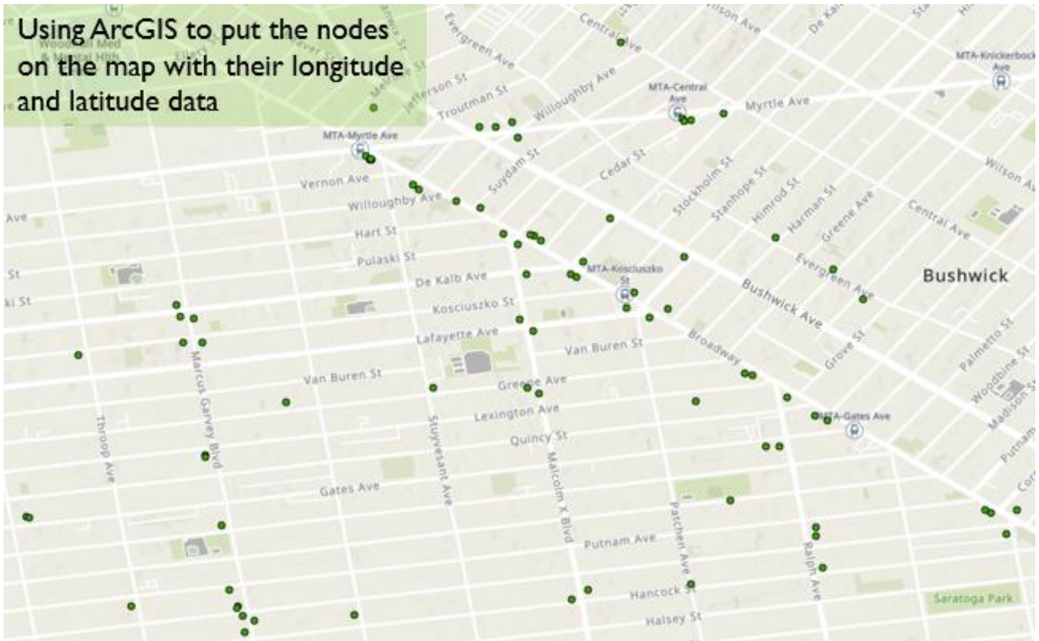

Figure 1. Network on the map (green dots: nodes, edges not shown)

3.1 Nodes

The nodes in this network are the restaurants, their coordinates are used to locate them on the map and their types (the types of foods they serve) will be used for further analysis. The total number of nodes is 124 so far, in the further research, the nodes represented awaiting areas will be added into the network.

3.2 Edges

The edges in this network are the paths connecting the restaurants to awaiting areas the driving distance (driving time) of the path will be the weight of each edge. The number of edges is unknown yet, however, it will be a multiple of the number of nodes(124), for the awaiting areas will be connected to all the restaurants in this region. Every combination of awaiting areas will increase the number of edges by 124.

4. Methodology

To look for the candidates' locations for awaiting areas, the team decided to use Kmeans clustering to find the candidates for this method can come out with points with higher closeness centralities, which means they will be closer to the other nodes. After observing the distribution of existed nodes in the network, the team decided to assign 3 awaiting areas for this region, also select 3 as the K value in Kmeans clustering.

4.1 Overall candidates by Kmeans

The coordinates of each restaurant are put into the algorithm, and the algorithm returned 3 candidate points in red circle. The coordinates of the overall candidates are: (-73.931 40.695), (-73.924 40.689), and (-73.940 40.688). The clustering plot is shown in Figure 2.

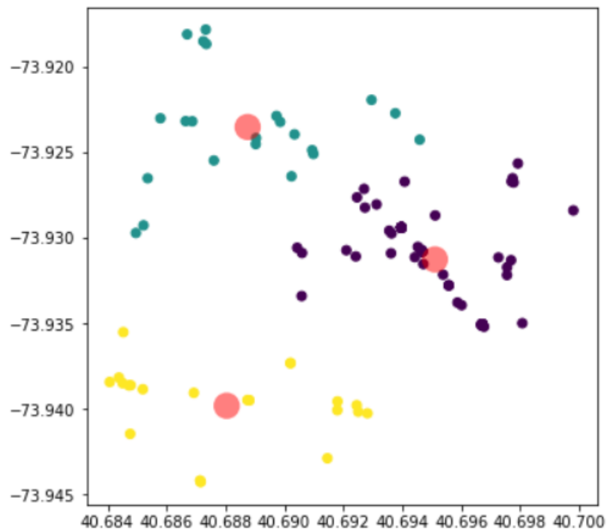

Figure 2. Overall candidate points prediction by Kmeans

4.2 Biased candidates by Kmeans

The overall clustering is obvious too simple and naïve. As some of the restaurants serves normal food, some of them serves desserts, smoothies or even groceries, the types of the restaurants are significantly different and they should not be considered on the same level.

4.2.1 Restaurant Classification

To make the clustering more reasonable, the team classified the restaurants into two categories, for the restaurants serving regular foods, they are classified as "major" group and there are 84 restaurants in this group, for the restaurants serving desserts, smoothies, or groceries, they are classified as "minor" group, and there are 40 restaurants in this group.

4.2.2 Clustering the Biased Groups Separately

The team then applied Kmeans clustering to these two groups, and got two sets of biased candidate points. Candidates by major group are (-73.923 40.689), (-73.940 40.688), and (-73.931 40.695), which are shown in Figure 3. candidates by minor group are (-73.941 40.690), (-73.925 40.689), and (-73.932 40.696), which are shown in Figure 4.

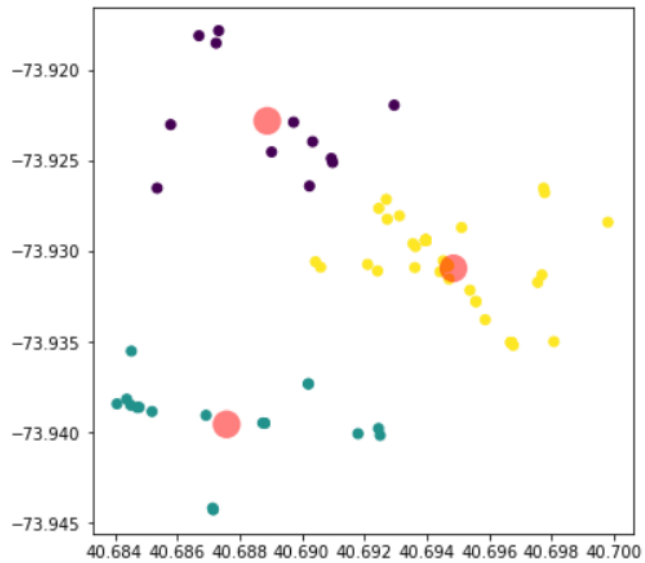

Figure 3. Major candidate points prediction by Kmeans

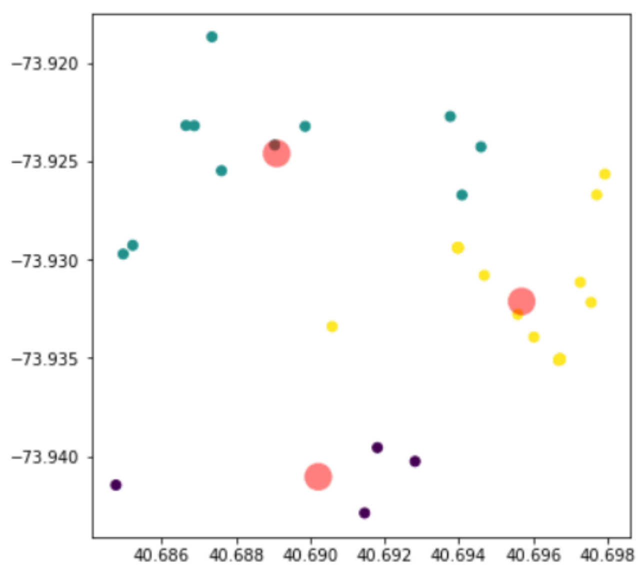

Figure 4. Minor candidate points prediction by Kmeans

4.3 Shifted Candidates Predicted by Biased Candidates

For the biased candidates suggesting the nodes with highest closeness in their groups, the candidates for all the restaurants should be located on the lines which connect the corresponding major and minor candidates.

4.3.1 Initial Candidates Predicted by Biased Candidates

If we ignore the occurrence of the orders, or say, the number of orders by each restaurant. The candidate points should be on the line but closer to the major candidates, for the major group has a larger number of restaurants than the minor group, which also means it is more likely to have orders by major restaurants. Being closer to the places where is more likely to have orders can shorten the response time for the delivers. However, the exact number of orders should be considered rather than the number of restaurants, for more restaurant does not necessarily get more orders proportionally. Additionally, the team thinks the deliver can pick up two orders from different groups if the time intervals of the order placements are shorter than two minutes, by picking up two orders in one route, the efficiency of delivery can be largely improved, and the team called this kind of orders as joint orders. When a joint order needs to be picked up, the best position to start the route should be at the middle of the major candidates and the minor candidates for the deliver can go any of the two restaurants as he/she wants first (any one which prepared the food faster).

4.3.2 Simulation to Determine the Shift Rates

In order to find out exact frequency of new orders from each group of restaurants, the team needs more specific data of orders by Doordash, however, it is not open to public due to the privacy of customers. The team decided to use two Poisson distributions to simulate the occurrence of joint orders. The team used the total number of review counts for all restaurants in a month to simulate the distribution, the numbers of orders will be proportional to the numbers of restaurants in the two groups. The Poisson distributions provides the team with the probabiliteis to have jonit orders within 2-min interval for each group (shown in Figure 5), and by adding two probabilites together, the team get the total probability (0.645) to have joint order.

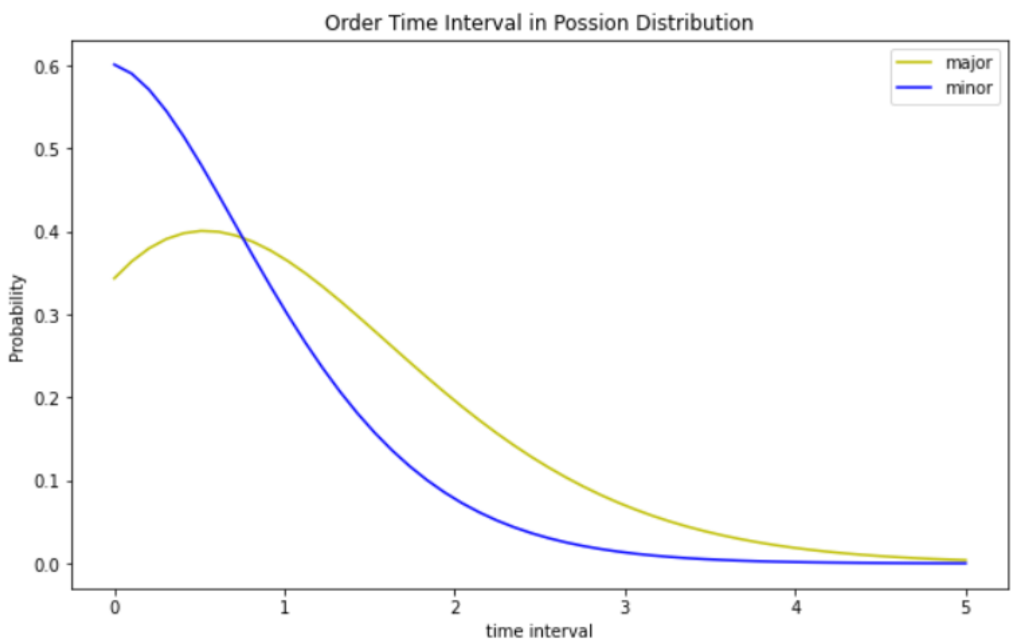

Figure 5. Joint orders' time intervals and its probabilities

As was discussed above, the existence of joint orders can "drag" the ideal candidates backwards the middle point of the connecting line of major and minor biased candidates from the initial candidates predicted by biased candidates. The team calculated the shifting rate of the initial candidates (In Equation 1).

$$Equation\ 1: Shifting\ Rate = \frac{84}{124} - \left[\left(\frac{84}{124} - \frac{1}{2}\right) \times 0.645\right] - 0.5 = 0.063$$

4.3.3 The Shifted Candidates

Considering the shifting effect brought by joint orders, the team use Equation 2 to improve the locations of initial candidates predicated by biased candidates.

$$Equation\ 2: Shifted\ Candidates = major\ candidates * 56.3\% + minor\ candidates * (1 - 56.3\%)$$

The coordinates of the shifted candidates are: (-73.940 40.689), (-73.931 40.695), and (-73.924 40.689).

4.4 Service Area Analysis on Candidates

The team then conducted service area analysis on the overall candidates and the shifted candidates using ArcGIS (shown in Figure 6).

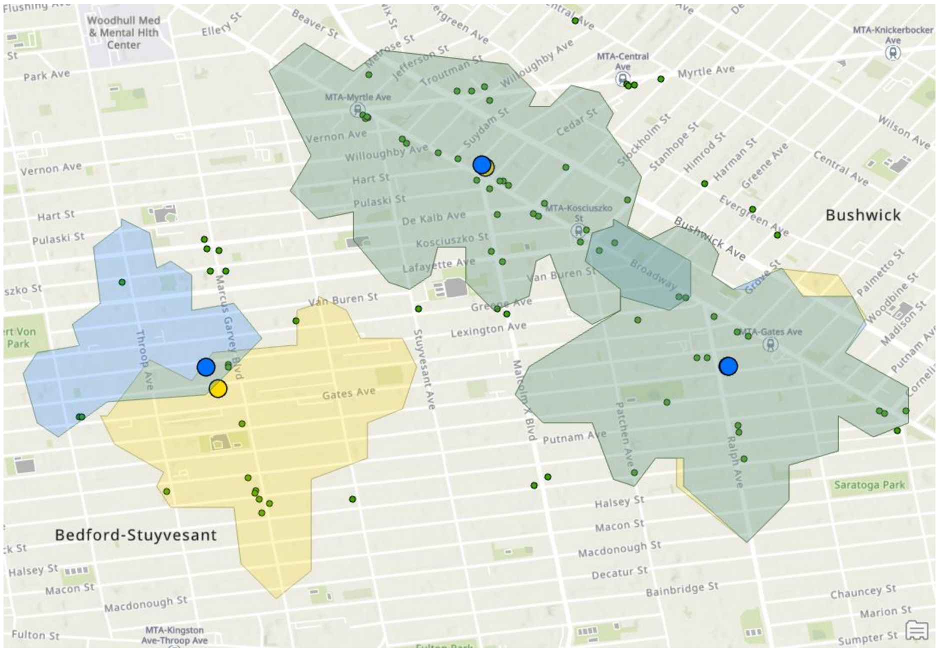

Figure 6. Service Area Analysis on overall candidates and shifted candidates

In the service area analysis, the yellow points represent the overall candidate points and the yellow areas represent the 2-min service area starting from overall candidates; the blue points represent the shifted candidate points and the blue areas represent the 2-min service area starting from shifted candidates; the green areas represent the 2-min service area covered by both sets of the candidate points.

5. Evaluation & Results

To decide whether using the overall candidates or the shifted candidates, the team counted the number of restaurants outside the 2-minutes service are for each case. For overall candidates, there are 23 restaurants outside its 2-minutes service area; for shifted candidates, there are 28 restaurants outside its 2-minutes service area. To cover more restaurant in 2-min service area, the team decided to choose overall candidate points as the final location to set up deliver awaiting areas, the coordinates of these three locations are (-73.931 40.695), (-73.924 40.689), and (-73.940 40.688).

6. Discussion & Potential Improvements

6.1 Discussion

The team did not expect that the overall candidates perform better than shifted candidates in the service area analysis. In Figure 5, the service areas of overall candidate and shifted candidate are significantly different, which is also unexpected, the team held the opinion that the existence one-ways in these blocks largely influence the service areas, the detours are not considered in the former algorithms.

6.2 Potential Improvements

In Kmeans clustering, the team used direct distances between two points to predict the candidate points, however, the real distances between two nodes should be the transportation distances (driving distances). In further research, the real distances should be put into the algorithm, which can help avoid the unexpected service area analysis results, for the detour can be considered as longer transportation distances in the algorithm.

Besides, the team used Poisson distribution to simulate the occurrence of orders, which is too simple compared to real-world circumstances. In further research, real order information should be used to predict the shift rate and determine the locations of candidate points.

7. Insights from the Lectures

7.1 Community Detection

Two community detection procedures are covered in this project, the first one is Kmeans clustering, it divided all restaurants into three groups and assign each group with one candidate point. The second detection is in the restaurant classification, as all the restaurants are divided into two groups by the kinds of food they serve.

7.2 Centrality

For the all the candidate points got by Kmeans clustering, they all have the highest closeness centralities in their group under their circumstance. Degree centralities are considered in this project, too. In the evaluation stage, the team counted the numbers of restaurant outside the service areas of the candidate points. Having more nodes in the service area means higher degree centrality, the candidates with higher degree centralities should be considered as better choices.

7.3 Shortest Path

In service area analysis, every path from the candidate points to the edge of their service area is the shortest path. In other words, ArcGIS’s algorithm uses shortest path to predict the largest area that can be covered in a certain period of time, in this case it is two minutes.

Reference

Y. Liu, B. Guo, C. Chen, H. Du, Z. Yu, D. Zhang, H. Ma, *FooDNet: Toward an Optimized Food Delivery Network Based on Spatial Crowdsourcing*, IEEE Transactions on Mobile Computing, vol. 18, no. 6, pp. 1288-1301, 1 June 2019, Doi: 10.1109/TMC.2018.2861864.

S. Li, L. Jiang, X. Wu, W. Han, D. Zhao, Z. Wang, *A weighted network community detection algorithm based on deep learning*, Applied Mathematics and Computation, Volume 401, 2021, 126012, ISSN 0096-3003, Retrieved from https://doi.org/10.1016/j.amc.2021.126012.