Operations Research Final Project
Optimal Planning for NTU YouBike Assignment

Group A

吳驊祐 B09705009 周成康 B09705011 陳瑾叡 B09705016
黃戎傳 B09705032 吳天冷 B08303135 李詠如 B07703084
劉德駿 B07801013 李宗霖 B07702044 陳琳瑄 B06B02059

June 10, 2022

1 Introduction

Many of us might have the following experience. After a tiring day at school, you plan to ride a YouBike to have a nice meal, however, there is no available YouBike in the station. After a long wait, you finally get a YouBike and arrive at the restaurant, you realize that there is no empty dock to return it. YouBike is supposed to be a convenience transportation option for citizens. However, such situation happens every day, and is kind of annoying.

This problem is due to the **imbalanced demand distributions**, which is related to our daily routine. For example, on weekday mornings, students ride YouBikes from MRT stations and dormitories to teaching buildings. We can observe that some stations receive YouBikes more than providing to others and some stations provide YouBikes more than receiving from others; therefore, YouBikes keep being accumulated in those "receiving-stations", while those "providing-stations" are consistently under-supplied. Although we often see the trucks shuttling back and forth in campus, working hard on repositioning YouBikes, the severe spatial imbalance still exists. Around NTU, there are **102 stations and 1800 docks** in total. During peak time, more than 300 YouBikes are needed to be repositioned while a truck can only ship 16 bikes at most. What's more, if we take a look at both large stations and small stations, 30% and 80% of docks are vacant, respectively. With the evidence mentioned above, this is indeed a problematic issue in urgent need of being solved.

Even though this kind of imbalance is inevitable and could not be entirely eliminated, the operators should minimize it, otherwise it could cause negative effects on revenue. Due to the reasons mentioned above, we decided to take a deep look into the issue, to see if we could figure out a better way to reposition YouBikes. There are only two options available, either the operators move YouBikes by themselves, or they provide customers incentives to do it for them (e.g., price discounts [1]).

[1]https://www.iot.gov.tw/dl-11486-61bf1c0d3d214be0b7800566fa19921d.html

In this project, we choose the former way and focus on the stations in NTU District. Although we focus on NTU District only, the same concept can be applied in any area. After all, to minimize the cost and maximize the efficiency, a single truck should not be in charge of a wide area. Suppose that the YouBike company decide to rent trucks for repositioning YouBikes and reschedule the routes. Our goal is to generate an optimal route that minimizes the cost and balances the percentage of available YouBikes in each station at the same time.

2 Problem Description

Our goal is to minimize the total cost and all possible negative impact, i.e., shortage and surplus that this dispatching problem may incur. This problem can be considered as a vehicle routing problem, which is NP-hard, and hence it is arduous to solve.

Interestingly, for this problem can be solved successfully, two virtual points are needed. Index 0 and index $N + 1$ both stand for virtual points, which are labeled as ”Start” and ”End”, respectively.

2.1 Decision Variables

2.1.1 Major Variables

The following variables are involved in the objective function, which means the objective value is determined by them.

y_{ij}	1 if the truck moves from station i to j and 0 otherwise, $y_{ij} \in \{0, 1\}$
w_{iL}	The percentage shortage after dispatching of station i, $w_{iL} \geq 0$
w_{iH}	The percentage surplus after dispatching of station i, $w_{iH} \geq 0$

Table 1: Major Decision Variables

Among all the major variables above, the y_{ij}'s are used for cost calculation. The w_{iL}'s and w_{iH}'s are related to penalty, if shortage and surplus happen to exist, i.e., the acceptable parking ratio is failed to be achieved, there should be sanctions imposed on the objective value.

2.1.2 Derived Variables

While the major variables directly determine the objective value, the derived variables are essential for constraints.

x_i	1 if station i visited by the truck and 0 otherwise, $x_i \in \{0, 1\}$
a_i	The number of bikes taken away from station i, $a_i \geq 0$
b_i	The number of bikes dispatched to station i, $b_i \geq 0$
p_i	The number of bikes on a truck when arriving at station i, $p_i \geq 0$
q_i	The number of bikes on a truck when leaving station i, $q_i \geq 0$
u_i	The order (number) in a truck's tour of station i, $u_i \geq 0$
v_{ij}	1 if station i is visited later than station j and 0 otherwise, $v_{ij} \in \{0, 1\}$

Table 2: Derived Decision Variables

Among all the derived variables above, the x_i's are crucial for restricting the y_{ij}'s. The a_i's, b_i's, p_i's and q_i's are required for the determination of the w_{iL}'s and w_{iH}'s. The u_i's and v_{ij}'s are critical for subtour elimination.

2.2 Parameters

Parameters Concerning Trucks and Bikes	
C_{ij}	The distance of a truck moving from station i to j
	(Let $C_{00} = C_{N+1,0} = C_{N+1,N+1} = 0$, $C_{0,N+1} = -T/G$.)
$Parked_i$	The number of bikes parked at station i at the moment
$Capacity_i$	The capacity of station i
M	The sum of parked bikes
G	The gas price per kilometer
T	The cost of renting a truck
Parameters Concerning Limitations	
A	The maximum capacity of a truck used for dispatching
L	The maximum distance for a truck to move
Parameters Concerning Penalty	
S_L	Lower bound of acceptable parking ratio
S_H	Upper bound of acceptable parking ratio
λ	The coefficient for penalty arising from shortage or surplus in dispatching
Parameters Concerning Subtours	
e	The constant for ensuring the relationship between the u_{ij}'s (orders).

Table 3: Parameters

It is worthwhile to notice that the S_L and S_H provide bounds for acceptable parking ratio so that an interval can be formed. The size of the interval should be correlated with λ, for the interval's size can be thought of as the "elasticity" for parking; different degrees of penalty should be granted to different elasticity. Therefore, we set the interval's size as the following.

$$[0.5 - \lambda, 0.5 + \lambda]$$

Note that the size of the interval is positively correlated with λ, in other words, larger interval implies that it is more inelastic to the number of bike changes and severer penalty should be meted out for not achieving the acceptable parking ratio.

2.3 Objective Function and Constraints

After defining our parameters and variables, the complete formulation can be constructed as the following.

2.3.1 Objective Function

The function can be split into three parts. The first term denotes the cost total cost of dispatching (traveling through stations); the second term represents the penalty associated with shortage and surplus, and the last term is the cost of renting a truck.

$$\min \quad G\sum_{i=0}^{N+1}\sum_{j=0}^{N+1} y_{ij}C_{ij} + \lambda(100\sum_{i=1}^{N}(w_{iL} + w_{iH}))^2 + T \tag{1}$$

2.3.2 Constraints for Routing

For the sake of moving from one station to another, both stations must be visited.

$$x_i + x_j \geq 2(y_{ij} + y_{ji}), \quad \forall i, j = 0, ..., N+1 \tag{2}$$

For each station, arrival implies departure, and vice versa.

$$\sum_{i=0}^{N} y_{ik} = \sum_{i=1}^{N+1} y_{kj}, \quad \forall k = 1, ..., N \tag{3}$$

It is not reasonable for a truck to move from a station to the same one.

$$y_{ij} = 0, \quad \forall i, j = 0, ..., N+1, i = j \tag{4}$$

A truck cannot move back to the previous station.

$$y_{ij} + y_{ji} \leq 1, \quad \forall i, j = 0, ..., N+1 \tag{5}$$

A truck cannot move for more than the maximum distance.

$$\sum_{i=0}^{N+1}\sum_{j=0}^{N+1} C_{ij}y_{ij} \leq L \tag{6}$$

A truck can only arrive at and leave a station once.

$$\sum_{i=0}^{N+1} y_{ij} \leq 1, \quad \forall j = 0, ..., N+1 \tag{7}$$

$$\sum_{j=0}^{N+1} y_{ij} \leq 1, \quad \forall i = 0, ..., N+1 \tag{8}$$

A visit of a station implies the arrival at it and / or the departure from it.

$$\sum_{j=0}^{N+1} y_{ij} + \sum_{j=0}^{N+1} y_{ji} \geq x_i, \quad \forall i = 0, ..., N+1 \tag{9}$$

2.3.3 Constraints for Bikes

Do not take away bikes and dispatch bikes at the same time.

$$a_i b_i = 0, \quad \forall i = 0, ..., N+1 \tag{10}$$

Take away / Dispatch bikes only if a station is visited.

$$a_i \leq M x_i, \quad \forall i = 0, ..., N+1 \tag{11}$$

$$b_i \leq M x_i, \quad \forall i = 0, ..., N+1 \tag{12}$$

Penalize when there is shortage or surplus.

$$w_{iL} \geq S_L - \frac{Parked_i - a_i + b_i}{Capacity_i}, \quad \forall i = 1, ..., N \tag{13}$$

$$w_{iL} \geq 0, \quad \forall i = 1, ..., N \tag{14}$$

$$w_{iH} \geq \frac{Parked_i - a_i + b_i}{Capacity_i} - S_H, \quad \forall i = 1, ..., N \tag{15}$$

$$w_{iH} \geq 0, \quad \forall i = 1, ..., N \tag{16}$$

The flow-in must be equivalent to the flow-out for each station.

$$\sum_{i=0}^{N+1} a_i x_i = \sum_{i=0}^{N+1} b_i x_i \tag{17}$$

Dispatch bikes to a station reasonably.

$$b_i + Parked_i \leq Capacity_i, \quad \forall i = 1, ..., N+1 \tag{18}$$

Take away bikes from a station reasonably.

$$a_i \leq Parked_i, \quad \forall i = 1, ..., N \tag{19}$$

2.3.4 Constraints for Trucks

When a truck visit a station, bikes are either taken away from or dispatched to it. The following equation describes the change in the number of bikes on a truck.

$$p_i + a_i - b_i = q_i, \quad \forall i = 0, ..., N+1 \tag{20}$$

The number of bikes on a truck when leaving a station must equal to that when arriving at the next one.

$$q_i y_{ij} = p_j y_{ij}, \quad \forall i, j = 0, ..., N+1 \tag{21}$$

The number of bikes on a truck cannot exceed the maximum capacity of it.

$$p_i \leq A, q_i \leq A, \quad \forall i = 0, ..., N+1 \tag{22}$$

The number of bikes dispatched to a station cannot be greater than that on a truck when arriving.

$$b_i \leq p_i, \quad \forall i = 0, ..., N+1 \tag{23}$$

2.3.5 Constraints for Virtual Points

No bikes are taken away from and dispatched to Start.

$$a_0 = 0, b_0 = 0 \tag{24}$$

No bikes are taken away from End, while it is acceptable to be dispatched to it.

$$a_{N+1} = 0, b_{N+1} \geq 0 \tag{25}$$

A truck must visit Start and End.

$$x_0 = 1, x_{N+1} = 1 \tag{26}$$

No arrival at Start, and no departure from End.

$$y_{i0} = 0, \quad \forall i = 1, ..., N+1 \tag{27}$$

$$y_{N+1,i} = 0, \quad \forall i = 0, ..., N \tag{28}$$

No bikes are on a truck at Start, and no bikes are leaving End. However, when arriving at End, the presence of bikes is acceptable.

$$p_0 = 0, q_0 = 0 \tag{29}$$

$$p_{N+1} \geq 0, q_{N+1} = 0 \tag{30}$$

No penalty for Start and End.

$$w_{0L} = 0, w_{0H} = 0 \tag{31}$$

$$w_{N+1,L} = 0, w_{N+1,H} = 0 \tag{32}$$

2.3.6 Constraints for Eliminating Subtours

In order to deal with subtours, the following constraints are indispensable.

$$
\begin{aligned}
&u_0 = 1 && && (33)\\
&u_i \leq N+2, && \forall i = 1, ..., N+1 && (34)\\
&u_i \geq 2, && \forall i = 1, ..., N+1 && (35)\\
&u_i - u_j + 1 \leq M(1 - y_{ij}), && \forall i, j = 0, ..., N+1, i \neq j && (36)\\
&u_i - u_j \leq M v_{ij} - e, && \forall i, j = 0, ..., N+1, i \neq j && (37)\\
&u_i - u_j \geq e - M(1 - v_{ij}), && \forall i, j = 0, ..., N+1, i \neq j && (38)
\end{aligned}
$$

2.3.7 Sign Constraints

$$
\begin{aligned}
&y_{ij} \in \{0, 1\}, && \forall i, j = 0, ..., N+1 && (39)\\
&w_{iL} \geq 0, w_{iH} \geq 0, && \forall i = 0, ..., N+1 && (40)\\
&x_i \in \{0, 1\}, && \forall i = 0, ..., N+1 && (41)\\
&a_i \geq 0, b_i \geq 0, p_i \geq 0, q_i \geq 0, && \forall i = 0, ..., N+1 && (42)\\
&u_i \geq 0, v_{ij} \in \{0, 1\}, && \forall i, j = 0, ..., N+1 && (43)
\end{aligned}
$$

The complete formulation is

$$
\begin{aligned}
&\min\ (1)\\
&\text{s.t.}(2) - -(43).
\end{aligned}
$$

3 Data Collection

The data is from Open Government Data Platform, which is a platform operated by National Development Council. The data are obtained by web crawling and then organized into the format shown as below.

	site_id	space_total	space_occupied	space_vacant	address	sarea	lati	long
0	0	37	0	37	汀洲路三段60巷2弄路側(A舍北側)	臺大專區	25.01493	121.53044
1	1	32	0	32	臺大水源舍區C南側	臺大專區	25.01466	121.52917
2	2	18	0	18	思源街16號之1旁	臺大專區	25.01411	121.52997
3	3	10	1	9	臺大檔案展示館東北側	臺大專區	25.01391	121.52895
4	4	30	7	23	汀洲路三段60巷2弄路側(B舍北側)	臺大專區	25.01525	121.53009
5	5	42	0	42	臺大男八舍東側	臺大專區	25.01729	121.54531

Figure 1: The Organized Data of YouBike 2.0

The latitude and longitude are used for calculating the distances among the stations with the haversine formula. However, using such method for distance calculation is evidently unrealistic in our research; therefore, Google Application Programming Interface (API) is applied instead so that distances for trucks to move can be calculated accurately.

4 Method

Our strategy is comprised of four procedures, which are **clustering, optimization, visualization, and analysis.**

First of all, k-`means` clustering algorithm is implemented for the purpose of dividing stations into several clusters, i.e., k groups. The clustering is based on the latitude-longitude coordinate system.

After that, different values of k and λ are chosen for invoking the **Gurobi solver** to optimize our problem, and thus different objective values with different combinations of k and λ are pinned down.

Consequently, different objective values reported by Gurobi are visualized so as to determine which combination of the two parameters to choose under a certain circumstance.

Lastly, the results acquired with the chosen pair of k and λ are further analyzed for our truck routing problem.

5 Results

We have tested our model with different data in different time to check feasibility. The following two instances demonstrate how the parameters are determined.

5.1 Instance 1: 6/1 17:00

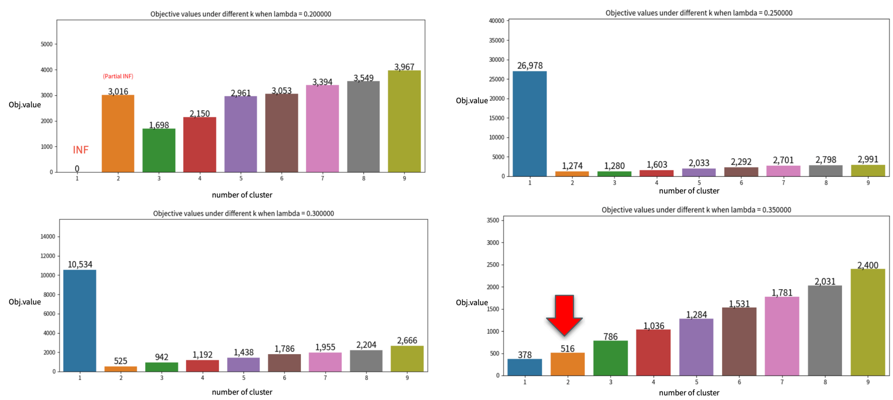

Figure 2: Objective Values under Different λs and Clusters in Instance 1

Figure 2 shows different objective values generated with different combinations of ks and λs. It is evident that larger λ leads to smaller objective value. Also note that some of the results in clustering is infeasible, where at least one group does not satisfy the time constraint, and therefore not considered in our decision making.

Given these four lambdas, if cost is our main consideration in decision making, the smallest λ and k, which are 0.35 and 2 respectively, would be chosen as our settings in parameters.

After λ and k are pinned down, our model would suggest routes for dispatching in figure 3. The model would also indicate where to start our assignment and either to take away bikes from or to dispatch them to a station. Before the assignment, the distribution of bikes is depicted as figure 4, where red spots denote the stations whose parking ratio lies beneath S_L while the blue ones represent the opposite, i.e., their parking ratios exceed S_H. After the assignment, we can see the results as figure 9 illustrates. It is apparent that either the number of red spots or that of blue spots decreases, which means that we have successfully move bikes to those stations that are under supplied and move out bikes from the over supplied.

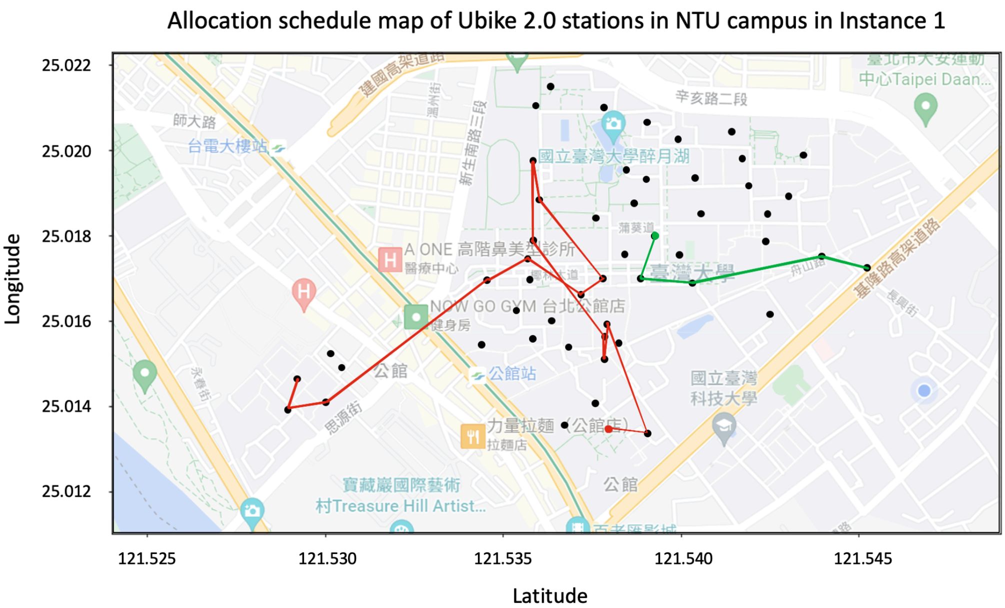

Figure 3: The Allocation Schedule Map in Instance 1

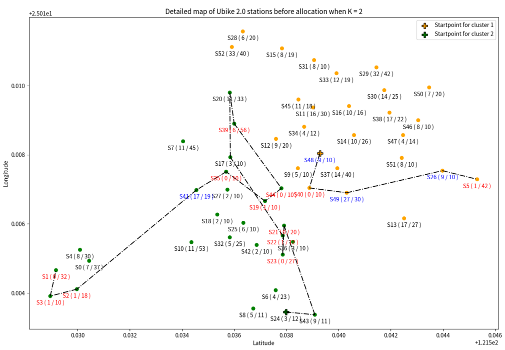

Figure 4: Detailed Ratio of YouBike Before Allocation in Instance 1

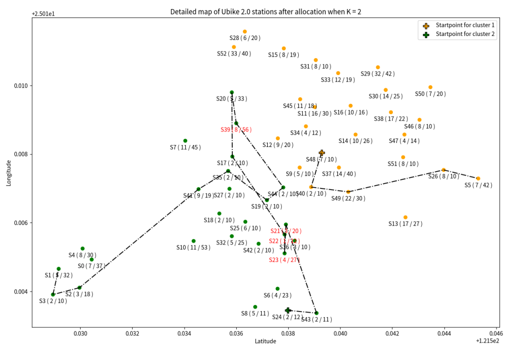

Figure 5: Detailed Ratio of YouBike After Allocation in Instance 1

5.2 Instance 2: 6/3 8:00

By same logic, that is, derive objective values under different combinations of λ and ks and then choose the appropriate pair of the parameters (as shown in figure 6). This time, λ is 0.35 and k is 8.

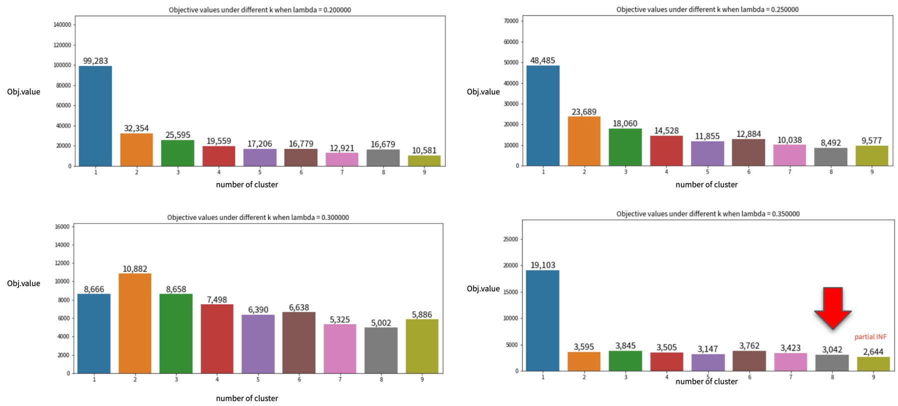

Figure 6: Objective Values under Different λs and Clusters in Instance 2

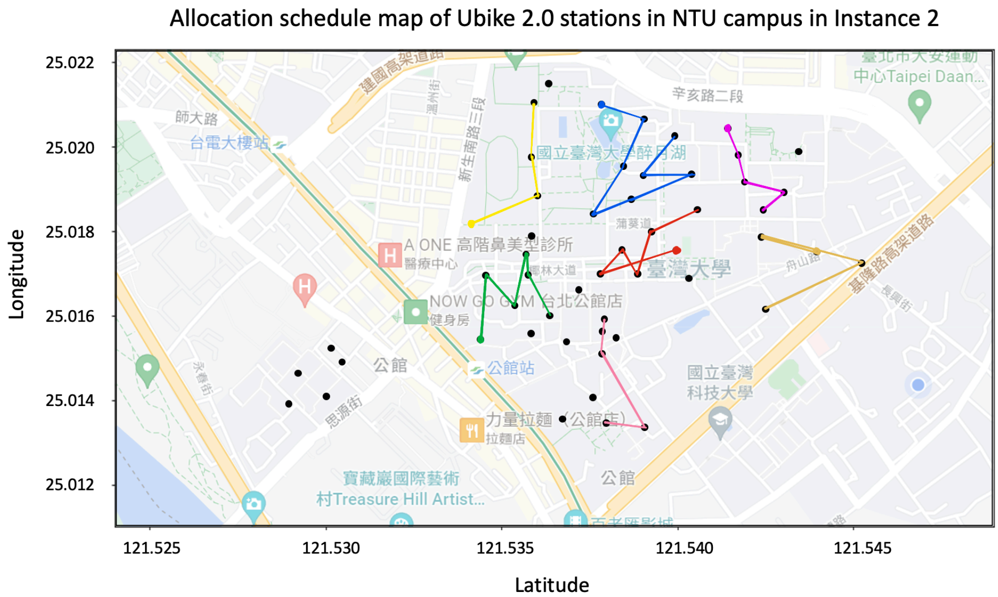

Figure 7: The Allocation Schedule Map in Instance 2

Figure 7 depicts the routing map when k is equal to 8. Notice that one of the groups (the bottom-left one) does not need dispatching. Before the assignment (as shown in figure 8), we may see that most of the stations are labeled in red, i.e., they are in shortages of bikes. Even though the assignment is completed (as shown in figure 9), the red spots are not all eliminated; there are still some of them existing.

Our inference is that the number of bikes in this certain time is not as much as usual, so even more trucks are assigned in dispatching, perfect distribution of bikes is still cannot be achieved.

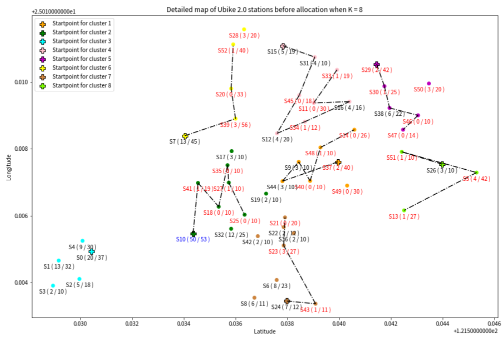

Figure 8: Detailed Ratio of YouBike Before Allocation in Instance 2

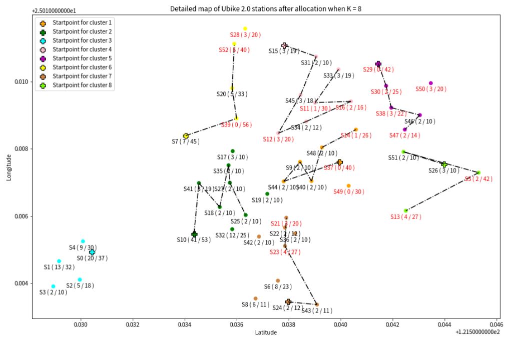

Figure 9: Detailed Ratio of YouBike After Allocation in Instance 2

What is worth mentioning is that larger λ implies larger interval size of acceptable parking ratio, which means that the restriction in parking ratio is "looser", and hence fewer bikes need to be dispatched. That is why in both instance 1 and 2, higher λ implies lower cost.

Therefore, to make a better decision is not focusing on picking out plans with lower cost adamantly. Choosing smaller λ under a certain budget constraint is a more flexible way.

6 Conclusions and Future Works

6.1 Conclusions

As mentioned previously, at the end the system lists the results of different (λ, k) and λ can be decided according to the operator's demand. For example, the four graphs in Figure 2 are all from instance 1 but with different λs. If the operator choose smaller λ, then the interval will be smaller. Therefore, it will be easier to violate the percentage constraint and increase the cost. On the other hand, if the operator choose larger λ, then the interval will become larger. The results might be the one whose situation cannot be improved significantly. In instance 1, the operator has to choose among 1274, 525 and 512. If there is enough budget, 1274 will always be recommended without a doubt. However, the operator usually needs to make a trade-off between cost and service. In summary, our system provides several suggestions, but the final decision still rests with the operator.

6.2 Future Works

6.2.1 Having More Than One Truck for Each Cluster

Because the process is dynamic, the number of clusters is not always the same. For example, in previous hour k is 4 while in this hour k is 2, then the rest of the two trucks can support other areas. In that case, a single cluster can have more than one truck.

6.2.2 Adding the Function for Data Collecting

As mentioned previously, the final decision still rests with the operator, although data collection is not in the field of operations research. However, by doing so, this system could assist the operator in a more comprehensive way.

6.2.3 Predicting the Demand Distribution for Next Period

So far, what we are doing is minimizing the imbalance already happened. In the next step, we can try to predict when the next peak time will be and what the demand distribution is by then. By doing so, we can reposition the YouBikes or move the redundant trucks for backup in advance.