Optimal Taxi Fleet Management

a Linear Programming Approach to the Taxi Capacity Problem

Jacky P.K. Li and Sandjai Bhulai

Vrije Universiteit Amsterdam,
Faculty of Science,
Amsterdam, The Netherlands
Email: {jacky.li, s.bhulai}@vu.nl

***Abstract*—This paper develops a model to determine the optimal number of taxis in a city by examining the trade-off between the overall profitability of the taxi service versus the customer satisfaction. We provide a data analytic investigation of taxi trips in New York City. We model the taxi service strategy by a fleet management model that can handle arrivals and deterministic travel times. Under this model, we examine the number of taxis in a particular period of time and measure the maximum profit in the overall system and the minimum number of rejected customer requests. We observe that the maximum profit of the overall system can be reduced significantly due to reducing the cost of driving without passenger(s). We present a case study with New York City Taxi data with several experimental evaluations of our model with a different period of time during the day and also with a realistic and a heuristic model. The results provide a better understanding of the requirement to satisfy the demand in a different period of time. These data may have important implications in the field of self-driving vehicles in the near future.**

Keywords–New York taxi service; revenue optimization; optimal routing; linear programming; min-cost network flow problem.

I. Introduction

Taxis are an essential component of the transportation system in most urban centers. The ability to optimize the efficiency of routing represents an opportunity to increase revenues for taxi services. Vacant taxis on the road waste fuel, represent uncompensated time for the taxi drivers, and create unnecessary carbon emissions while also generating additional traffic in the city. In the not-too-distant future, fully autonomous vehicles will be the norm rather than the exception. Taxis could eventually work together to satisfy the demand of the customers versus compete against each other to make revenue individually. This can reduce the amount of traffic on the road and the overall fuel cost significantly. Based on these ideas, creating a model in which all the taxis work together to satisfy all the customers would be an interesting endeavor to explore the number of taxis necessary to satisfy all the demand.

Previous studies have focused on developing recommendation systems for taxi drivers [1]–[6]. Several studies use the global positioning system (GPS) to create recommendations for both the drivers and the passengers to increase profit margins and reduce seek times [3] [5]–[7]. Ge et al. [8] and Ziebart et al. [9] gather a variety of information to generate a behavioral model to improve driving predictions. Ge et al. [1] and Tseng et al. [10] measure the energy consumption before finding the next passenger. Castro et al. [7], Altshuler et al. [11], Chawla et al. [12], Huang et al. [13], and Qian et al. [14] learn knowledge from taxi data for other types of recommendation scenarios, such as fast routing, ride-sharing, or fair recommendations.

In terms of Linear Programming research, Liang et al. [15] propose a method of automated vehicle operation in taxi systems that addresses the problem of associating trips to automated taxis; however, this research paper is based on a small case study. It does not provide a feasible model. Roling et al. [16] describe an ongoing research effort pertaining to the development of a surface traffic automation system that will help controllers to better coordinate surface traffic movements related to arrival and departure traffic in airport traffic planning. Bsaybes et al. [17] developed framework models and algorithms for managing a fleet of Individual Public Autonomous Vehicles with a heuristic model.

In this paper, we are confronting the problem within the context of managing New York City (NYC) taxis to serve customers who request a ride. We are investigating a realistic model with 15 km x 15 km grid combined with a demand of 20,000 rides in 30 minutes. We assume the total business time is equal to the sum of the total occupancy time plus the total seeking time. Fundamentally, if we can satisfy the ride requests with deterministic travel times and minimize the seeking time, this would provide the maximum profit in the overall system.

The deterministic version of this problem is the min-cost/max-profit integer problem. The linear and integer versions for the min-cost "multi-commodity-flow" problem have been studied extensively in [18] and [19]. In this paper, we also examine both a realistic and a heuristic model to explore the difference between the two models with real New York City Taxi data that is provided to the public. Figure 1 shows that there are consistent patterns in demand between certain periods of the day and certain days of the week during June 2013.

The paper is structured as follows. In Section II, we analyze

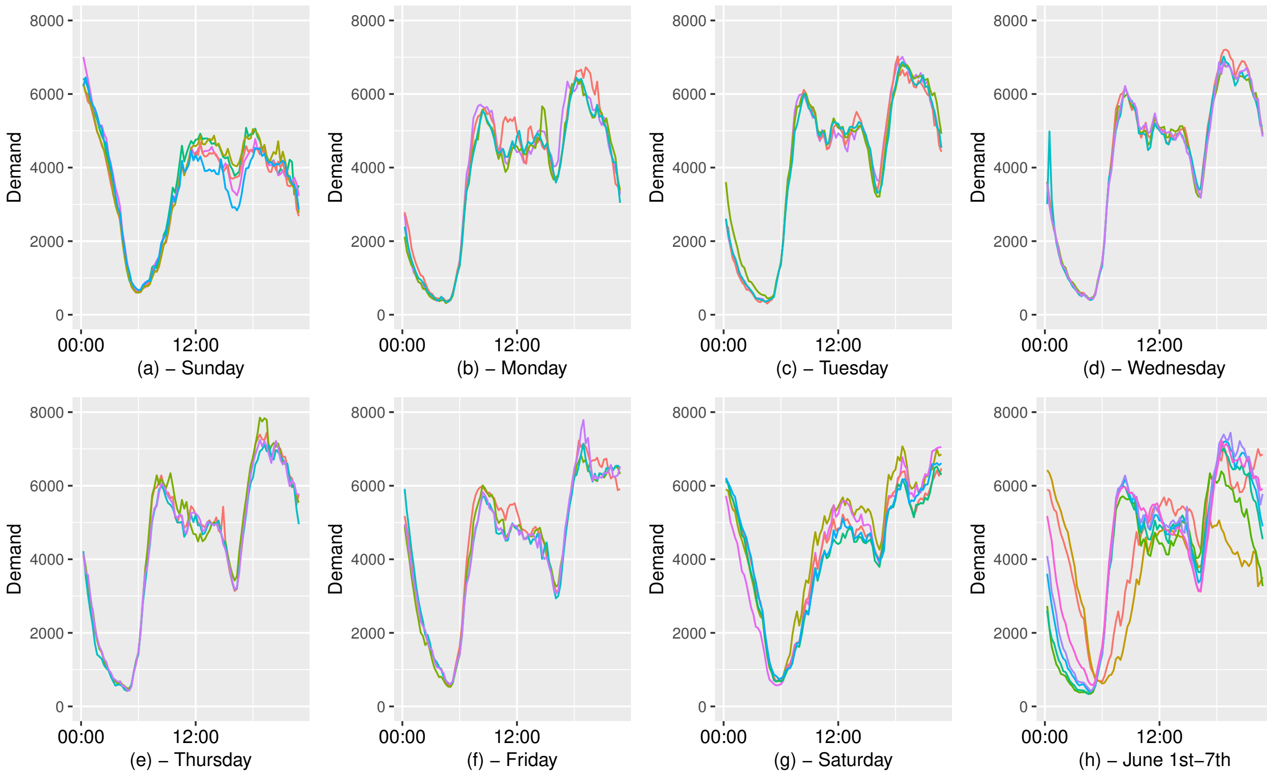

Figure 1. (a) to (h) display the day of the week in June. (i) displays the first week of June.

the New York Taxi dataset in 2013. This provides the input for our linear programming model, which is explained in Section III. We assess the performance of the linear programming model in Section IV, where we conduct numerical experiments with the realistic model and also the heuristic model. Finally, the paper is concluded in Section V.

II. Dataset

In our research, we are investigating NYC taxi demand patterns of a particular day of the week. From each ride record, we use the following fields: pick-up time, pick-up longitude, pick-up latitude, drop-off time, drop-off longitude, drop-off latitude, and traveling time. We omit the records containing missing or erroneous GPS coordinates. Records that represent trip durations longer than 1 hour and trip distances greater than 100 kilometers are omitted.

We are interested in observing a consistent demand during a period of a month. According to timeanddate.com [20], only three days in the month of June recorded a rainfall. We believe the weather and temperature can be a factor of the demand. Figure 1 displays the days of the week in June from Sunday (Figure 1a) to Saturday (Figure 1g), respectively, and also the first week of June in Figure 1h.

During the weekday, the lowest demand of the day is from approximately 04:00 to 05:00, and the highest demand is from approximately 18:00 to 19:00 followed by 08:00 to 08:30 and the 12:00 to 12:30 period. Based on this observation, we choose the four different time slots, the lowest demand of the day 04:00-04:30, the morning traffic 08:00-08:30, the lunch break 12:00-12:30, and the dinner traffic 18:30-19:00. Table I displays the maximum, the average, the minimum demand and the coefficient of variation per minute during four different time periods in June. The coefficient of variation is consistent especially for Tuesday and Wednesday. Based on this observation, we choose June 4th, 2013, the first Tuesday of June as our main focus.

The analysis of the dataset of the New York Taxi service is focused on the island of the Manhattan area in New York, USA. This area imposes a rectangular grid of avenues and streets. We discretized the grid into a 50×50 grid, making each block in the grid approximately 300 meters $\times$ 300 meters. The choice for a block size of 300 meters is based on the assumption that a taxi can traverse this distance within 1 minute. Due to the calculation time, we also created a heuristic model with a 10×10 grid, making each block in the grid approximately 1,500 meters $\times$ 1,500 meters and a taxi can traverse this distance within 5 minutes.

The state of a taxi can be described by two parameters: the current location, which is an element of the set $L = \{(1,1),\ldots,(50,50)\}$ grid and the current time, which comes from the set $T = \{1,\ldots,30\}$. We will denote the system state in our model as $s = ((i,j),t) = (i,j,t)$, which we will elaborate on in Section III.

We select June 4th, 2013 as the date to analyze with an average of 36.47 requests per minute from 04:00-04:30, an average of 581.37 requests per minute from 08:00 to 08:30, an average of 472.43 requests per minute from 12:00 to 12:30, and an average of 661.90 requests per minute from 18:30 to

TABLE I. THE TOTAL DEMAND PER MINUTE IN JUNE 2013.

		Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
04:00-04:30	Max	314	63	56	72	80	116	304
	Average	233.22	39.22	35	49.21	56.86	83.56	55.95
	Min	145	20	18	31	40	53	102
	% Coefficient of variation	16.50%	18.75%	19.48%	15.45%	16.41%	11.17%	28.09%
08:00-08:30	Max	174	588	631	643	679	637	284
	Average	130.07	516.43	562.88	567.71	575.06	539.53	186.95
	Min	80	418	502	506	477	439	134
	% Coefficient of variation	14.44%	7.42%	5.51%	5.48%	6.96%	7.42%	7.20%
12:00-12:30	Max	485	562	551	560	568	570	586
	Average	435.21	462.97	495.86	502.38	501	489.47	477.51
	Min	382	389	419	441	420	432	399
	% Coefficient of variation	6.10%	8.33%	5.26%	4.17%	6.90%	7.06%	8.06%
18:30-19:00	Max	526	686	760	735	739	709	722
	Min	380	552	563	559	575	565	497
	Average	456.81	622.58	654.07	642.47	664.57	637.41	593.03
	% Coefficient of variation	7.37%	4.33%	5.56%	5.78%	5.34%	5.57%	7.53%

19:00.

III. LINEAR PROGRAMMING

The deterministic version of the taxi routing problem is by solving a max-profit integer "multi-commodity-flow" problem for each time period. These problems tend to get large easily with the number of possible states and resource types, and their multi-commodity nature presenting an unwelcome dimension of complexity. Due to this reason, we present both a realistic and a heuristic model for comparison.

For notational convenience, we denote (x, y) by i and denote (x', y') by j. We are required to serve every customer demand. However, if there are not enough vacant taxis within the same grid, the unsatisfied customer demands are not served. To handle this, we assume that the unsatisfied demands are lost, and we take the profit from serving a higher revenue demand to be the incremental profit from serving the demand with a taxi.

In [21], we assumed that all taxis take a single time period to travel and all customers have the same taxi preferences. In this model, we extend our formulation to cover cases where there are multi-period travel times. For notational convenience, we assume that demand at a certain location can be served by a taxi at the same location at the same time and the demand can be served by the vacant taxi that is able to arrive at the same location at the same time. For the rest of the section, we adopt the terminology that an empty taxi driving toward the next customer is "seeking".

A. Parameters of the Linear Programming Model

In this subsection, we state the parameters used in the rest of the model.

Location – realistic $(i, j) \in L = \{1, \ldots, 50\} \times \{1, \ldots, 50\}$: the area is divided into a grid of 50×50 grid cells;

Location – heuristic $(i, j) \in L = \{1, \ldots, 10\} \times \{1, \ldots, 10\}$: the area is divided into a grid of 10×10 grid cells; we implement a smaller grid compared to the realistic model in order to simplify the calculation process.

Time $t \in T = \{1, \ldots, 30\}$: we use minutes as the interval of a time slot, and a total of 30 minutes as time horizon.

- $D_{i,j,t}$ describes the number of **demand** that need to be carried from grid cell i to grid cell j at time period t from the original dataset on June 4th, 2013.

- $S_{i,j,t}$ describes the number of **empty** taxis moving from grid cell i to grid cell j at time period t from the original dataset.

- $\mathcal{T}_{i,j,t}$ describes the **traveling time** from a taxi moving from grid cell i to grid cell j. We assume the traveling time is the same at any period of time t. In the heuristic model, the **traveling time** is multiplied by 5 to match the travel time for both models since the grid size is also increased by a factor of 5.

- $x^{\mathrm{l}}_{i,j,t}$ describes the number of **loaded** taxis moving from grid cell i to grid cell j at time period t.

- $x^{\mathrm{e}}_{i,j,t}$ describes the number of **empty** taxis moving from grid cell i to grid cell j at time period t.

- $c^{\mathrm{l}}_{i,j}$ describes the net **reward** from an occupied taxi moving from grid cell i to grid cell j. We assume the profit is the same at any period of time t. In the heuristic model, the $c^{\mathrm{l}}_{i,j}$ is multiplied by 5 to match the realistic model.

- $c^{\mathrm{e}}_{i,j}$ describes the **cost** of a vacant taxi moving empty from grid cell i to grid cell j. We assume the cost is the same at any period of time t. (Remark: In order to simplify the model, the **cost** is half of the **reward** and the heuristic model is multiplied by 5 to match the realistic model.)

- $R_{i,t,t'}$ describes the number of taxis in operation that are inbound to location i at time period t and will arrive at location i at time period t'.

- $\mathcal{R}$ describes the number of taxis in the system.

The deterministic version of the problem we are interested in can be written as:

$$\max \sum_{t \in T} \sum_{i,j \in L} (-c^{\mathrm{e}}_{i,j} x^{\mathrm{e}}_{i,j,t} + c^{\mathrm{l}}_{i,j} x^{\mathrm{l}}_{i,j,t}) \tag{1}$$

TABLE II. DEMAND AND SEEKING ON JUNE 4TH, 2013

	Realistic (50×50)					Heuristic (10×10)				
	Actual Demand	# of Vehicles	Demand / # of Vehicles	Driving empty	Missed demand	Actual Demand	# of Vehicles	Demand / # of Vehicles	Driving empty	Missed demand
04:00-04:30	1,094	650	1.68	628	10	1,035	600	1.73	240	16
08:00-08:30	17,441	8,000	2.18	5,469	7	15,982	7,350	2.17	2,185	30
12:00-12:30	14,173	5,400	2.62	5,213	11	12,800	5,050	2.53	1,699	36
18:30-19:00	19,857	8,200	2.42	5,015	12	18,203	7,650	2.38	1,124	13

subject to

$$
\begin{aligned}
&R_{i,1,t'} = \mathcal{R}, && i \in L, t \in \{1\}, t' \in T,\\
&\sum_{j\in L}(x^{\mathrm{e}}_{i,j,t} + x^{\mathrm{l}}_{i,j,t}) = R_{i,t',t}, && i \in L, t, t' \in T,\\
&R_{j,t',t+1} = \sum_{i\in L} I_{(t'-t)=\tau_{i,j}}(x^{\mathrm{e}}_{i,j,t} - x^{l}_{i,j,t}) + R_{j,t',t}, && \\
& && j \in L, t, t' \in T,\\
&x^{\mathrm{l}}_{i,j,t} \le D_{i,j,t}, && i, j \in L, t \in T,\\
&x^{\mathrm{e}}_{i,j,t}, x^{\mathrm{l}}_{i,j,t} \in \mathbb{Z}_+, && i, j \in L, t \in T.
\end{aligned}
\tag{2}
$$

which is a special case of the max-profit integer multi-commodity flow problem.

We evaluate the linear programming approach based on the New York Taxi dataset on June 4th, 2013 on four particular times 04:00-04:30, 08:00-08:30, 12:00-12:30, and 18:30-19:00. In our deterministic case study experiment, we formulate the problem as a max-profit integer problem (1). From the dataset, we generate the data of $D_{i,j,t}$, which is the number of demand from location i to location j at time t. We also generate the data of $S_{i,j,t}$, which is the number of empty taxis driving from location i to location j at time t to seek for the next passenger(s).

In order to provide a better understanding of our result, we calculate:

- Demand $= D_{i,j,t}$,
- Actual seeking $= S_{i,j,t}$,
- Seeking from our model $= c^{\mathrm{e}}_{i,j}$,
- Missing demand $= D_{i,j,t} - x^{*,\mathrm{l}}_{i,j,t}$,
- Actual revenue $= c^{\mathrm{l}}_{i,j} \times D_{i,j,t}$,
- Revenue from our model $= c^{\mathrm{l}}_{i,j} \times x^{*,\mathrm{l}}_{i,j,t}$,
- Actual cost $= c^{\mathrm{e}}_{i,j} \times S_{i,j,t}$,
- Cost from our model $= c^{\mathrm{e}}_{i,j} \times x^{*,\mathrm{e}}_{i,j,t}$,
- Actual profit = actual revenue $-$ actual cost,
- Profit = revenue $-$ cost,
- Lost revenue $= c^{\mathrm{l}}_{i,j} \times [D_{i,j,t} - x^{*,\mathrm{l}}_{i,j,t}]$,

where $x^{*,\mathrm{e}}_{i,j,t}$ and $x^{*,\mathrm{l}}_{i,j,t}$ are the optimal solutions for $x^{\mathrm{e}}_{i,j,t}$ and $x^{\mathrm{l}}_{i,j,t}$, respectively.

One thing to note is that the initial location of the vehicle was set up based on $R_{i,1,t'} = \mathcal{R}$, which means the vehicles are located to the highest demand profit at the start.

IV. CASE STUDIES AND OBSERVATIONS

In this section, we concentrate on the programming and the results. We use `Rcplex` [22] to run our model. In the realistic model we calculate the result based on the 50×50 grid and a 30-minute interval. Due to the constraints and variables, the matrix size is approximately 377 million x 188 million. This requires an approximate 250 GB of memory according to [22], it took over three hours per calculation. This model is do-able in terms of calculation, but not scalable making it intractable for larger problem sizes. Due to this reason, we create the heuristic model with 10×10 grid to decrease the size such that it can be handled on a standard desktop computer with 8 GB of memory. We also increase the travel time in the heuristic model by 5 to match the increase in size of the grid. The calculation time with the heuristic model is approximately 10 seconds.

Table II displays the results of four different time periods using both the realistic and the heuristic model. The heuristic model has less demand due to having no demand within the same grid which eliminates just under 10% of the requests each time period. The demand for the number of vehicles is close in each model under both the realistic and the heuristic models. These results indicate that the simplified heuristic model can still give an accurate approximation of how many taxis we need to satisfy the demand per period of time.

Table III displays the results of each minute for both the realistic and the heuristic model for 12:00-12:30 on June 4th, 2013. We decrease the size of the fleet by 1,000 vehicles and see the percentage difference of the profit. The optimal fleet sizes are 5,400 for the realistic model and 5,050 for the heuristic model. When we increase the fleet size by a 1,000, there will be less driving without the passenger(s) and it does not provide more profit to the system.

Figure 2 provides a view of the profit comparison for the realistic model on June 4th, 2013 from 12:00 to 12:30. A fleet of 5,000 vehicles would satisfy most of the demand and provide the highest demand in the system.

V. CONCLUSION

We use a linear programming to model the taxi service and determine the optimal policy to yield the best profit in the overall system. In Table II, each taxi can cover approximately 1.76, 2.17, 2.60, and 2.40 demand at 04:00-04:30, 08:00-08:30, 12:00-12:30 and 18:30-19:00 time periods, respectively, for both models.

Table III displays the profit for both the realistic and the heuristic model. The optimal solution for the realistic model requires 5,400 vehicles which provide an increased profit of 25,844.50 units to the actual profit and is unable to satisfy

TABLE III. TABLE OF DEMAND, SEEKING, ACTUAL PROFIT AND DIFFERENCE OF PROFIT WITH DIFFERENT SIZE OF THE FLEET FOR 12:00-12:30 TIME PERIOD FOR BOTH THE REALISTIC AND THE HEURISTIC MODEL TO COMPARE BETWEEN THEM.

	Realistic Model (50×50)						Heuristic Model (10×10)					
Minute	Demand $D_{i,j,t}$	Seeking $S_{i,j,t}$	Actual Profit	$6,400$ Vehicles	$5,400$ Vehicles	$4,400$ Vehicles	Demand $D_{i,j,t}$	Seeking $S_{i,j,t}$	Actual Profit	$6,050$ Vehicles	$5,050$ Vehicles	$4,050$ Vehicles
1	464	409	3,826.5	+917.5	+917.5	+917.5	415	251	3,725	+935	+935	+920
2	490	416	3,951.5	1,021	1,021	1,021	439	277	3,830	+1,060	+1,060	+1,060
3	476	411	4,110	+934.5	+934.5	+934.5	433	255	3,977.5	+932.5	+932.5	+922.5
4	484	374	4,139	+842	+842.5	+842.5	441	219	4,125	+810	+810	+810
5	461	373	3,819	+920.5	+923	+923	417	217	3,892.5	+892.5	+892.5	+892.5
6	462	385	3,777	+956	+959	+959.5	418	243	3,747.5	+962.5	+972.5	+870
7	471	401	4,012	+968	+971	+970.5	423	246	3,902.5	+957.5	+960	+925
8	475	369	4,212.5	+819	+824	+816.5	434	207	4,122.5	+777.5	+777.5	+697.5
9	491	363	4,483.5	+769	+791	+790	456	213	4,532.5	+752.5	+752.5	+722.5
10	479	341	4,224	+828.5	+850.5	+831.5	438	198	4,292.5	+777.5	+777.5	+682.5
11	470	372	4,088.5	+896	+918	+877.5	434	215	4,052.5	+822.5	+820	+622.5
12	441	367	3,479	+882	+915	+871.5	393	221	3,350	+850	+895	+790
13	418	359	3,373.5	+871.5	+923	+789	378	201	3,397.5	+797.5	+857.5	+632.5
14	453	353	3,766	+825	+873	+638.5	406	198	3,662.5	+757.5	+772.5	+410
15	490	326	4,409	+725.5	+777.5	469	454	193	4,365	+695	+690	+340
16	505	342	4,752	+762	+818.5	+414	463	206	4,727.5	+747.5	+767.5	-217.5
17	482	346	4,324.5	+809.5	+865	+485.5	436	201	4,360	+750	+767.5	-97.5
18	500	390	4,206.5	+905.5	+964	+287.5	453	228	4,212.5	+862.5	+902.5	+2.5
19	493	372	4,164	+825.5	+866	+189.5	439	211	4,120	+800	+912.5	-55
20	490	342	4,079.5	+776	+831	+79.5	431	195	3,887.5	+707.5	+737.5	+52.5
21	453	394	3,569.5	+913.5	+996	+422.5	398	220	3,497.5	+797.5	+860	+165
22	437	359	3,758	+813	+892	+322	397	201	3,792.5	+702.5	+800	+167.5
23	504	360	4,509.5	+790.5	+876	+332.5	450	202	4,442.5	+717.5	+747.5	-105
24	479	335	4,262	+746	+827	+158	438	184	4,190	+645	+717.5	+85
25	458	332	3,593	+789.5	+861	+231.5	392	196	3,387.5	+737.5	+822.5	+527.5
26	469	301	4,232.5	+657	+747.5	+240.5	410	167	4065	+605	+600	+170
27	504	319	4,804	+694.5	+759	+298.5	461	179	4715	+635	+625	+50
28	440	344	4,116	+719.5	+768	+516	404	194	4115	+660	+655	+350
29	466	337	3,830	+739.5	+740	+498	433	201	3837.5	+702.5	+687.5	+227.5
30	468	299	3,988	+593	+593	+489	416	179	3942.5	+622.5	+622.5	+402.5
Total	14,173	10,791	121,860	+24,710.5	+25,844.5	+17,616.5	12,800	6,318	120,267.5	+23,472.5	+24,130	+13,022.5
Missing Demand	-	-	-	0	11	1,469	-	-	-	0	36	139

11 demands. In the heuristic model, it requires 5,050 vehicles to satisfy 12,800 demands and it misses 36 demands in that period. The table also shows that more vehicles to satisfy all the demand does not provide the highest profit.

As for future discussion, our current model is based on a 50 $\times$ 50 grid with a 30-minute time period, which is 2,500 $\times$ 2,500 $\times$ 30 = 375 million data points on one dimension, this requires a supercomputer with 2 TB memory and it takes over three hours per calculation. This is not a feasible method to solve the model. Creating a model that uses dynamic programming with value function approximation will reduce the calculation time and the memory use.

Secondly, having stochastic demand would provide an even more realistic model, especially when traffic accidents occur in real time. Lastly, ride sharing is an obvious next step toward taxi routing research. Can we satisfy all the demand with limited vehicles and maximize the profit?

REFERENCES

[1] Y. Ge et al., "An energy-efficient mobile recommender system," in Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '10. New York, New York, USA: ACM Press, 2010, p. 899. [Online]. Available: http://dl.acm.org/citation.cfm?doid=1835804.1835918

[2] H. Rong, X. Zhou, C. Yang, Z. Shafiq, and A. Liu, "The rich and the poor: A Markov decision process approach to optimizing taxi driver revenue efficiency," in Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. ACM, 2016, pp. 2329–2334.

[3] J. Yuan, Y. Zheng, L. Zhang, X. Xie, and G. Sun, "Where to find my next passenger," in Proceedings of the 13th international conference on Ubiquitous computing. ACM, 2011, pp. 109–118.

[4] Y. Zheng, J. Yuan, W. Xie, X. Xie, and G. Sun, "Drive Smartly as a Taxi Driver," in 2010 7th International Conference on Ubiquitous Intelligence & Computing and 7th International Conference on Autonomic & Trusted Computing. IEEE, Oct 2010, pp. 484–486.

[5] M. Qu, H. Zhu, J. Liu, G. Liu, and H. Xiong, "A cost-effective recommender system for taxi drivers," in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14. New York, New York, USA: ACM Press, 2014, pp. 45–54.

[6] D. Zhang et al., "Understanding Taxi Service Strategies From Taxi GPS Traces," IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 1, pp. 123–135, Feb 2015.

[7] P. Castro, D. Zhang, and S. Li, "Urban traffic modelling and prediction using large scale taxi GPS traces," Pervasive Computing, pp. 57–72, 2012.

[8] Y. Ge, C. Liu, H. Xiong, and J. Chen, "A taxi business intelligence system," in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '11. New York, New York, USA: ACM Press, 2011, p. 735. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2020408.2020523

[9] B. D. Ziebart, A. L. Maas, A. K. Dey, and J. A. Bagnell, "Navigate like a cabbie: Probabilistic reasoning from observed context-aware behavior," in Proceedings of the 10th international conference on Ubiquitous computing. ACM, 2008, pp. 322–331.

[10] C.-M. Tseng and C.-K. Chau, "Viability analysis of electric taxis using New York city dataset," in Proceedings of the Eighth International Conference on Future Energy Systems. ACM, 2017, pp. 328–333.

[11] T. Altshuler, R. Katoshevski, and Y. Shiftan, "Ride sharing and dynamic networks analysis," arXiv preprint arXiv:1706.00581, 2017.

[12] S. Chawla, Y. Zheng, and J. Hu, "Inferring the Root Cause in Road Traffic Anomalies," in 2012 IEEE 12th International Conference on Data Mining. IEEE, Dec 2012, pp. 141–150.

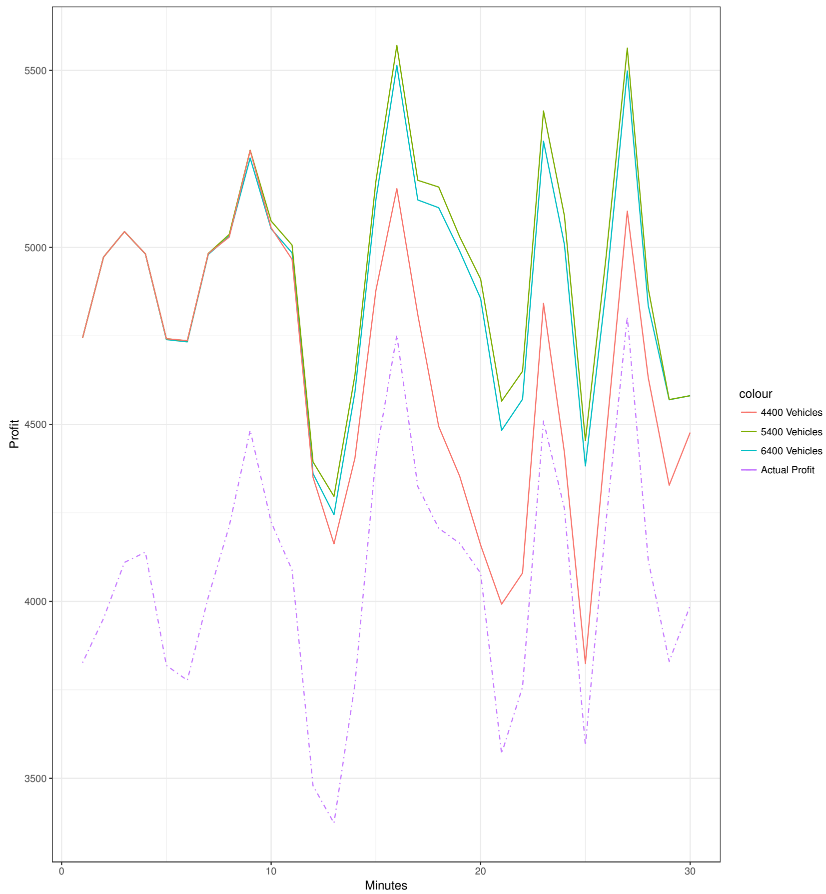

Figure 2. Profit with the different sizes of the vehicles inventory in 30 minutes.

[13] Y. Huang, F. Bastani, R. Jin, and X. S. Wang, "Large scale real-time ridesharing with service guarantee on road networks," Proceedings of the VLDB Endowment, vol. 7, no. 14, pp. 2017–2028, 2014.

[14] S. Qian, J. Cao, F. L. Mouël, I. Sahel, and M. Li, "Scram: a sharing considered route assignment mechanism for fair taxi route recommendations," in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015, pp. 955–964.

[15] X. Liang, G. H. de Almeida Correia, and B. van Arem, "An optimization model for vehicle routing of automated taxi trips with dynamic travel times," Transportation Research Procedia, vol. 27, pp. 736–743, 2017.

[16] P. C. Roling and H. G. Visser, "Optimal airport surface traffic planning using mixed-integer linear programming," International Journal of Aerospace Engineering, vol. 2008, no. 1, p. 1, 2008.

[17] S. Bsaybes, A. Quilliot, and A. K. Wagler, "Fleet management for autonomous vehicles using multicommodity coupled flows in time-expanded networks," in LIPIcs-Leibniz International Proceedings in Informatics, vol. 103. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[18] R. Mesa-Arango and S. V. Ukkusuri, "Minimum cost flow problem formulation for the static vehicle allocation problem with stochastic lane demand in truckload strategic planning," Transportmetrica A: Transport Science, vol. 13, no. 10, pp. 893–914, 2017.

[19] D.-P. Song and J. Carter, "Optimal empty vehicle redistribution for hub-and-spoke transportation systems," Naval Research Logistics (NRL), vol. 55, no. 2, pp. 156–171, 2008.

[20] CustomWeather. (2013) Past weather in New York, New York, USA June 2013. [Online]. Available: https://www.timeanddate.com/weather/usa/new-york/historic?month=6&year=2013

[21] J. Li, S. Bhulai, and T. van Essen, "Dynamic coordination of the New York city taxi to optimize the revenue of the taxi service," to appear in International Journal On Advances in Intelligent Systems, 2018.

[22] Technote. (2012) Guidelines for estimating CPLEX memory requirements based on problem size. [Online]. Available: https://www-01.ibm.com/support/docview.wss?uid=swg21399933