October 1, 2009

A Service Choice Model for Optimizing Taxi Service Delivery

Shih-Fen Cheng, *Singapore Management University*
Xin Qu

A Service Choice Model for Optimizing Taxi Service Delivery

Shih-Fen Cheng
School of Information Systems
Singapore Management University
sfcheng@smu.edu.sg

Xin Qu
School of Information Systems
Singapore Management University
xinqu@smu.edu.sg

***Abstract*— Taxi service has undergone radical revamp in recent years. In particular, significant investments in communication system and GPS devices have improved quality of taxi services through better dispatches. In this paper, we propose to leverage on such infrastructure and build a service choice model that helps individual drivers in deciding whether to serve a specific taxi stand or not. We demonstrate the value of our model by applying it to a real-world scenario. We also highlight interesting new potential approaches that could significantly improve the quality of taxi services.**

I. Introduction

Taxi service is of significant importance in metropolitan areas, however, even after decades of studies and improvements, its operational efficiency is still not quite satisfactory (e.g., a taxi fleet would spend over 50% of time idling in a typical day). The reason behind this inefficiency is how the taxi service is organized: in most cities, taxi services are delivered either by pre-arranged pick-ups (e.g., Dial-a-Cab service in UK and Singapore), street pick-ups, or taxi stand pick-ups. Although modern taxi-dispatch systems (e.g., see [1] and [2]) can satisfy pre-arranged pick-up requests very efficiently, significant portion of demands still have to be delivered by the latter two choices, which are highly dynamic and variable. This is one of the main reasons why taxi services are still delivered very inefficiently.

To counter this uncertainty and come up with sensible strategy, taxi drivers have to rely on their own experiences and usually very limited real-time information. In other words, most services (those not pre-arranged) are delivered by drivers using local heuristics without global information. Thus, if we take a bird's-eye view over the urban area, there are always imbalances of supplies and demands in different sub-areas. Some areas might have far more taxis than potential demand while some areas might experience shortage in taxis and a surge in customer waiting time. The implication of this is that simply by reducing imbalances, the quality of service could be improved without policy interventions (e.g., increase the quota for taxi licenses or modify taxi fares).

In this paper, we present a simple queuing model for making optimal servicing decisions for individual taxis at a major fixed location (e.g., airports, train stations, or tourist attractions). Important parameters required for solving the model are assumed to be available through Global Position System (GPS) that is installed on each individual taxi. To infer model parameters from the noisy and sometimes inaccurate GPS signals, we also propose a practical filtering and post-processing procedure. The applicability of our methodology is demonstrated by a real-world scenario.

The paper is organized as follows. In section II, we describe the problem and its background, some related works are also discussed. In section III, we formally introduce the service choice model for making service decision at a busy taxi stand. In section IV, we discuss how to incorporate GPS traces into the abstract model we just described. In section V, we present a real-world application based on GPS data obtained through our partner. Finally, we conclude the paper in section VI.

II. Background

Taxi service has long been studied in the queuing and transportation literature. Kendall [3] proposed to model taxi service delivery at a taxi stand as a *double queue* in which both customers and taxis form separate queues. Sasieni [4] later extended the model such that either customers or taxis might leave the queue if they have waited too long. In more recent literature, models have also been proposed for taxis that roam the street in providing services (e.g., see [5] and [6]).

Despite these past efforts in coming up with models for different modes of taxi operations, it is still not clear how one can integrate these models to obtain an unified model for making optimal servicing decisions. In other words, we need to build a model when taxi drivers could freely switch servicing modes between roaming the street, waiting for dispatch, or serving a taxi stand.

In the next section, we would propose one such integrated model that considers two operational modes: roaming the traffic network and serving a taxi stand.

III. The Service Choice Model

As mentioned in the previous section, we attempt to create a model that includes multiple operational modes. More specifically, we assume that taxi drivers can either choose to serve the general traffic network by roaming, or alternatively, they can choose to wait at one particular taxi stand.

Generally speaking, serving a taxi stand could provide more stable income since demand pattern at a particular taxi stand is usually more recurrent and predictable than roaming on the road. However, the benefit of serving the taxi stand would vanish quickly with the length of the queue.

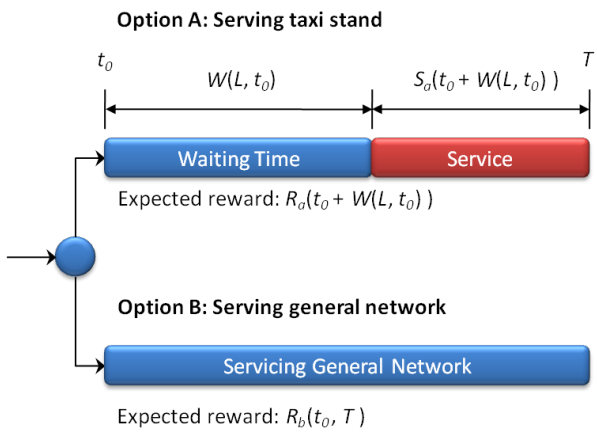


Fig. 1. The service choice model with two options.

Following this intuition, if we could characterize a function that takes queue length as input and returns relative benefits for serving the taxi stand, then the optimal operational policy should be threshold based. This threshold length can then be easily used by taxi drivers or taxi operators to efficiently identify locations that are *over served* and *under served*.

The service choice model is illustrated in Figure 1. Following the earlier stated assumption, there are two options available in our service choice model: (a) serve the taxi stand, and (b) serve the general network. The rewards for serving both options are denoted as $R_a(\cdot)$ and $R_b(\cdot)$ respectively.

For each taxi that chooses option A at time t_0 and observes queue length L, a waiting time $W(L, t_0)$ is expected. At the end of the waiting period, a customer will be served, with expected revenue $R_a(t_0 + W(L, t_0))$ and expected service time $S_a(t_0 + W(L, t_0))$.

For option B, on the other hand, the $R_b(\cdot)$ can be viewed as the opportunity cost for choosing option A. Therefore, $R_b(.)$ can be computed as the expected revenue from time t to time T. T is the expected service termination time for choosing option A. Suppose we can compute average time-dependent operation revenue for the general traffic network, and denote it as $r(t)$, we can then compute $R_b(\cdot)$ as:

$$R_b(t, T) = \int_{t_0}^{T} r(t)dt.$$

Without loss of generality, assume that $W(.)$ is a non-decreasing function in L. We can then obtain the threshold policy by finding the largest L that satisfies $R_a(\cdot) \geq R_b(\cdot)$. This problem can be formally defined as an optimization problem:

$$\begin{aligned} \max \quad & L \qquad (1) \\ \text{s.t.} \quad & \\ & R_a(t_0 + W(L, t_0)) \geq R_b(t, T) \\ & L \geq 0 \end{aligned}$$

In some cases, Problem (1) might have no solution. This simply implies that option B dominates option A, and serving taxi stand is never recommended.

To solve Problem (1) properly, we need expressions for $R_a(\cdot)$, $R_b(\cdot)$, and $W(\cdot)$. Since these expressions are usually scenario dependent, we will defer the discussion of them to section V, when we explore a real-world case study.

IV. Incorporating GPS Traces

In section III, a service choice model is proposed to provide a conceptual framework in reasoning about the optimal policy within a taxi service network. A critical missing piece in the model, as stressed in section III, is the lack of information for characterizing important model data.

One potential source for supplying this information is the GPS devices that are quite commonly used in modern taxi fleets. These devices could provide information regarding real-time taxi locations as well as paid trips. Unfortunately, there are at least two issues we need to resolve before using these GPS traces:

1) Civilian GPS devices usually produce significant errors. To use them as sources of information, we need to perform proper filtering.
2) The information provided by GPS traces is factual snapshots of real-time taxi service network. However, we could not obtain information like waiting time or queue length directly. These information needs to be inferred by other models.

To handle the first issue, we adopt a filter that could infer network topology from GPS traces and subsequently drop traces that are unrelated to the specific taxi stand. The handling of the second issue is based on the same filtering procedure: after all traces are marked to the network topology, the boundary of the queue can be identified and the queue build-up and the waiting time can then be detected accordingly.

The filter procedure is best illustrated with the numerical example. The technique will be described in detail in subsection V-B.

V. Case Study

A. Case Background

To illustrate how the service choice model could be implemented in the real-world scenario, we collaborate with one major taxi service provider in Singapore and acquire its operational data. This taxi company is the dominant player in Singapore, with over 60% of market share and a fleet size of more than 15,000 taxis. Each and every taxi in it fleet is equipped with a specially-designed mobile data terminal (MDT) that embeds GPS and allows bi-directional text communications. The MDT is designed to provide basic GPS functionalities and transmit back (to the central server) current vehicle positions every 30 to 60 seconds; for each paid trip, all trip-related information is also sent back and recorded on the central server (e.g., type of job, starting

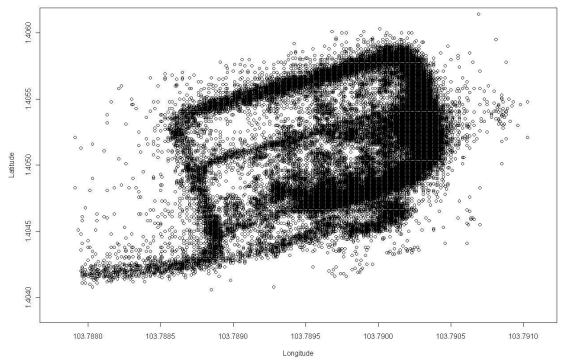


Fig. 2. The superposition of all GPS traces from September, 2008.

Fig. 3. The satellite map of the Night Safari area (taken from Google Maps).

and ending location, total fare). This data set thus allows us to perform the study on service choice at virtually any location in Singapore.

Since the focus of this paper is on how to make optimal decisions at a taxi stand, we pick a popular yet relatively isolated tourist spot so that unrelated traffic and noises could be minimized. The location we choose to study is the Night Safari, which is one of the most popular tourist attractions in Singapore. Besides geographically isolated, the Night Safari is also shown to be an ideal location for study since it only opens from 7pm to midnight and its volume of visitors is quite stable.

In this study, we use data collected from the month of September in 2008. In this data set, there are 95,390 GPS traces and they are plotted according to their coordinates in Figure 2. From Figure 2 we could clearly see that despite our best effort in choosing an isolated location, the selected data set is still very noisy, and we need to properly filter it before performing any meaningful analysis.

B. Processing Noisy GPS Traces

As introduced in section IV, by building the filtering procedure, we attempt to achieve two goals: 1) remove noisy GPS traces, and 2) infer unobservable yet important information like queue length and queuing time.

1) Learning Network Topology: The first step in building our filtering procedure is to learn the network topology from the GPS traces[1]. Despite the noisiness of GPS traces in Figure 2, we can still easily recognize the road network topology underneath these traces. In fact, by comparing the plot of traces to the map of the same region in Figure 3, we can see that the majority of these traces actually falls within reasonable range from the actual road. This observation motivates an intuitive yet effective filtering algorithm that is based on the well-known k-means clustering algorithm. Bruntrup et al. [7] have proposed a similar clustering approach for map generation, however, their focus is on map generation, not subsequent queuing analysis.

In general, the k-means clustering will divide points into k clusters (k is pre-determined) such that the sum of the squared distance between each point and the center of the cluster it belongs to would be minimized. The k-means algorithm is an iterative heuristic that could approximate the clustering. The algorithm proceeds as follows:

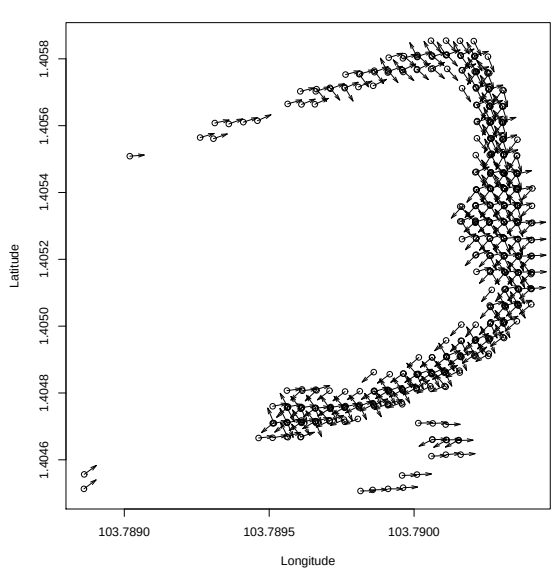
1) Determine the number of clusters, k.
2) Randomly generate k clusters and determine the cluster centers (pick from one of the points), or directly generate k points as cluster centers.
3) For each point that is not clustered yet, assign it to the closest cluster center.
4) After all points are clustered, update centers of all clusters.
5) Repeat steps 3 and 4 until converge (or when changes in centers as within certain threshold).

Although this is an approximation algorithm, it scales quite well and works very efficiently in practice. It should be noted that the final clustering would heavily depend on the initial clustering and might not be global optimal clustering.

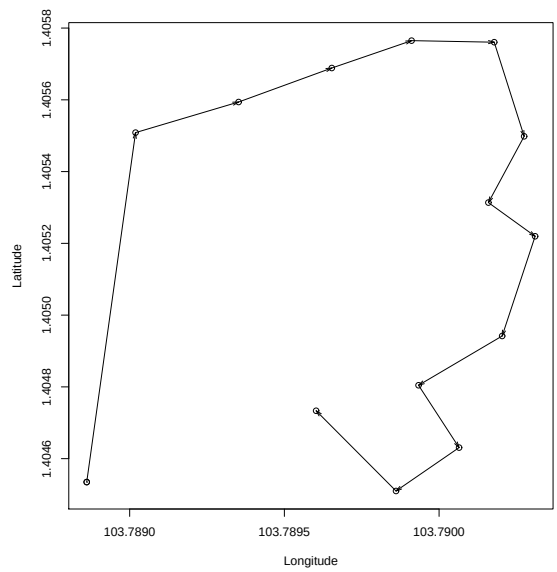
For each GPS trace in our application, it is defined by three parameters: longitude (x), latitude (y), and direction (θ). x and y are readings directly available from GPS, while θ is obtained by first connecting a sequence of traces produced by the same vehicle and then measuring the angle from the connecting arc to the horizontal line in counterclockwise direction, as illustrated in Figure 4.

In our numerical analysis, we perform two phases of k-means clustering; in the first phase we let $k = 220$, and with these 220 cluster centers, a second phase is executed with $k = 14$. From our empirical experience, we find that such multi-phase design allows better and smoother path aggregation. The progression of the algorithm is illustrated in Figure 5.

[1]The purpose of this step is to ensure that we could correctly identify routes taken by queuing taxis even when these routes are not on the digital map.



(a) 220 cluster centers after the first phase.



(b) 14 cluster centers after the second phase, with connection applied.

Fig. 5. The two phases of the k-means clustering algorithm.

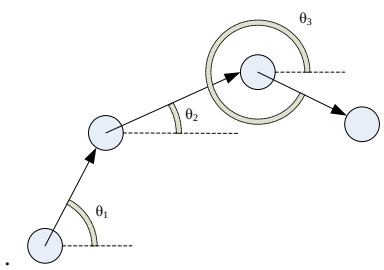


Fig. 4. Examples on how θ's can be computed for a sequence of GPS traces.

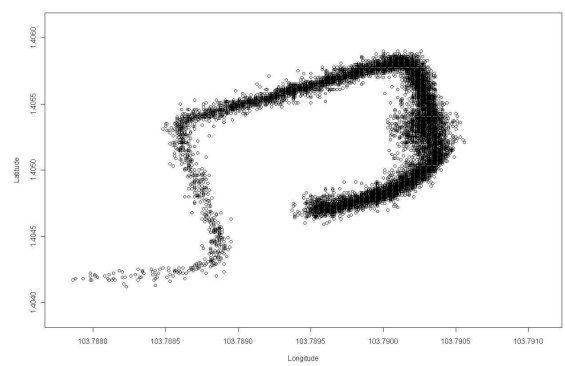


Fig. 6. The remaining GPS traces after filtering.

To infer the network topology from these clusters, we would define a distance measure between any pair of clusters i and j to be:

$$S_{ij} = \alpha\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} + \beta|\theta_i - \theta_j|.$$

By specifying a starting cluster i_0, we could find the next cluster to connect to by finding $j = \arg\min_j S_{i_0,j}$. By repeating this step iteratively, we could then identify a chain of cluster centers which can be viewed as the base network topology that characterizes the queuing area for the taxi stand.

2) Filtering GPS Traces: After the network topology is learned, it can then be used for filtering GPS traces. For each GPS trace, we first try to identify the road segment that is closest to it. This road segment is then regarded as the base road segment for this GPS trace and the distance from the trace point to the road is recorded in a *distance set* associated with this road segment.

After distance sets are created, we can then construct the confidence intervals around each road segment. In our analysis, all GPS traces that fall out of the 90% confidence intervals around any road segment are discarded. In Figure 6 we can see the remaining GPS traces after filtering.

For each taxi, by connecting valid traces belong to it, its queuing route can then be constructed. Here we would perform the final checking: the order in which road segments appear in a taxi's route should match the chain of clusters. The taxi trip should be dropped if it fails this final test.

C. Identifying Queue Length and Waiting Time

Two most important pieces of information, queue length and waiting time, can also be inferred by using the learned network topology. Based on the learned network topology, we define the first node as the entering point of the queue and the last node as the exiting point of the queue.

TABLE I

OPTIMAL POLICIES FOR DIFFERENT TIME PERIODS

Time periods	19:00-20:00	20:00-21:00	21:00-22:00	22:00-23:00	23:00-24:00
Avg. customer number (per hour)	4.3	13.8	31.0	31.6	14.1
Avg. fare from Night Safari: $R_a(t)$	15.22	15.16	14.78	15.22	17.57
Avg. service time from Night Safari: $S_a(t)$ (min)	21.61	22.79	22.25	21.47	20.93
Avg. customer inter-arrival time: $m(t)$ (min)	14.61	4.50	2.01	1.97	4.42
Avg. general revenue per hour: $r(t)$	20.08	19.23	18.85	18.76	16.10
Threshold Policy (L^*)	0.83	2.57	10.56	15.79	23.96
Avg. queue length	2.3	7.3	14.6	18.4	13.4
Avg. waiting time (min)	32.5	36.4	31.2	42.6	22.3

If we view the GPS traces we receive from the central server as discrete events, then whenever a taxi crosses the entering point to enter the queue, the queue counter will be increased by one and the entering time will be recorded. Similarly, whenever a taxi crosses the exiting point to exit the queue, the queue counter will be decreased by one and the exiting time will be recorded. For this taxi, its waiting time is simply the difference between the exiting time and the entering time.

D. Optimal Policy and Analysis

Before deriving the optimal policy for our service choice model, we need to first define $R_a(\cdot)$, $S_a(\cdot)$, $R_b(\cdot)$, and $W(\cdot)$. $R_a(t)$ and $S_a(t)$ is defined as the expected revenue and service time *per trip* out of Night Safari at time t. $R_b(t_1, t_2)$, on the other hand, is defined as the expected revenue during the period $[t_1, t_2]$, and is simply computed by $R_b(t_1, t_2) = r(t_1) \cdot (t_2 - t_1)$, where $r(t_1)$ is expected revenue *per hour* at time t_1.

Finally, for $W(L, t)$, we assume it to be a linear function, in which the service time for each taxi is a time-dependent constant (i.e., the expected intervals in between customer arrivals). Therefore, $W(L, t) = m(t) \cdot L$, in which $m(t)$ denotes the customer's inter-arrival time at time t.

With these function definitions, L^* can be found by solving:

$$R_a(t) = \frac{S_a(t) + W(L^*, t)}{60} \cdot r(t).$$

By substituting functions according to the above definitions, we have:

$$L^* = \frac{1}{m(t)} \left[\frac{60 \cdot R_a(t)}{r(t)} - S_a(t) \right].$$

In our analysis, we define each time period to be one-hour long and compute relevant constants accordingly. The summary of the analysis can be seen in Table I. Note that for the last time period (23:00 to 24:00), the policy we computed is 23.96, however, the expected number of customer in that hour is only 14.1. This seemingly unreasonable is obtained because we used inter-arrival time in computing L^*. To use such policy in realistic setting, one should also consider expected *remaining* demand in the time period. Suppose the chance of seeing additional demand is low (e.g., number of customer shown up has already exceeded the average), taxi drivers should be advised not to join the queue. In other words, L^* should be capped by the expected remaining demand, and L^* in the last time period should thus be reduced to 14.1, the same as expected demand.

To contrast how real drivers perform against the optimal threshold policy we obtain, we also include the average queue length drivers observe at the point of entry and the average waiting time. As shown in Table I, drivers tend to over-serve the Night Safari, particularly during the busier hours.

Another interesting finding is that the average waiting time during the busiest hour (22:00-23:00) is actually the longest. This might be counter-intuitive at first sight, but it makes sense after we realize that drivers are ill-informed on the queuing situations and also other's intention. When the only information drivers know is that 22:00-23:00 is the most profitable time period, all self-interested drivers will simultaneously decide to go to the Night Safari if they can, and as game theorist would predict, the globally inefficient joint actions would prevail exactly because everyone is selfish.

VI. CONCLUSION

In this paper we propose how to derive an optimal service choice model for taxi drivers serving a taxi stand. By incorporating GPS data, this model can be used in real-time and assist drivers in improving the delivery of their services (and in the process, earn more income). We demonstrate the potential of our methodology by applying our model to a real-world scenario. As expected, most drivers, when making their decisions independent, are making sub-optimal decisions.

We believe that the performance of overall taxi fleet can be significantly improved if proper mechanism could be introduced so that optimal decisions computed by our model can be revealed to appropriate drivers. This area, along with the real-world implementation of our approach, are our major future research directions.

VII. ACKNOWLEDGEMENT

The authors gratefully acknowledge support from ComfortDelGro in providing its operational data. The authors also would like to thank C. Jason Woodard and Darshan

Santani for their valuable suggestions and advices in the early stage of this work. This work was partially supported by the School of Information Systems, Singapore Management University.

REFERENCES

[1] Z. Liao, "Real-time taxi dispatching using global positioning systems," *Communications of the ACM*, vol. 46, no. 5, pp. 81–83, 2003.

[2] D.-H. Lee, H. Wang, R. L. Cheu, and S. H. Teo, "A taxi dispatch system based on current demands and real-time traffic information," *Transportation Research Record*, vol. 1882, pp. 193–200, 2004.

[3] D. G. Kendall, "Some problems in the theory of queues," *Journal of Royal Statistics Society B*, vol. 13, no. 2, 1951.

[4] M. W. Sasieni, "Double queues and impatient customers with an application to inventory theory," *Operations Research*, vol. 9, no. 6, pp. 771–781, 1961.

[5] H. Yang and S. C. Wong, "A network model of urban taxi services," *Transportation Research Part B: Methodological*, vol. 32, no. 4, pp. 235–246, 1998.

[6] K. I. Wong, S. C. Wong, and H. Yang, "Modeling urban taxi services in congested road networks with elastic demand," *Transportation Research Part B: Methodological*, vol. 35, no. 9, pp. 819–842, 2001.

[7] R. Bruntrup, S. Edelkamp, S. Jabbar, and B. Scholz, "Incremental map generation with gps traces," in *IEEE Conference on Intelligent Transportation Systems*, 2005.