Towards Reducing Taxicab Cruising Time Using Spatio-Temporal Profitability Maps

Conference Paper · August 2011

DOI: 10.1007/978-3-642-22922-0_15 · Source: DBLP

4 authors, including:

Jason W. Powell
University of North Texas
4 PUBLICATIONS **38** CITATIONS
SEE PROFILE

Yan Huang
University of North Texas
85 PUBLICATIONS **1,757** CITATIONS
SEE PROFILE

Minhe Ji
East China Normal University
23 PUBLICATIONS **223** CITATIONS
SEE PROFILE

Towards Reducing Taxicab Cruising Time Using Spatio-Temporal Profitability Maps

Jason W. Powell[1], Yan Huang[1], Favyen Bastani[1], and Minhe Ji[2]

[1] University of North Texas
{jason.powell, huangyan**}@unt.edu, FavyenBastani@my.unt.edu
[2] East China Normal University
mhji@geo.ecnu.edu.cn

Abstract. Taxicab service plays a vital role in public transportation by offering passengers quick personalized destination service in a semi-private and secure manner. Taxicabs cruise the road network looking for a fare at designated taxi stands or alongside the streets. However, this service is often inefficient due to a low ratio of *live miles* (miles with a fare) to *cruising miles* (miles without a fare). The unpredictable nature of passengers and destinations make efficient systematic routing a challenge. With higher fuel costs and decreasing budgets, pressure mounts on taxicab drivers who directly derive their income from fares and spend anywhere from 35-60 percent of their time cruising the road network for these fares. Therefore, the goal of this paper is to reduce the number of cruising miles while increasing the number of live miles, thus increasing profitability, without systematic routing. This paper presents a simple yet practical method for reducing cruising miles by suggesting profitable locations to taxicab drivers. The concept uses the same principle that a taxicab driver uses: follow your experience. In our approach, historical data serves as experience and a derived Spatio-Temporal Profitability (STP) map guides cruising taxicabs. We claim that the STP map is useful in guiding for better profitability and validate this by showing a positive correlation between the cruising profitability score based on the STP map and the actual profitability of the taxicab drivers. Experiments using a large Shanghai taxi GPS data set demonstrate the effectiveness of the proposed method.

Key words: Profitability, Spatial, Temporal, Spatio-temporal, Taxi, Taxicabs

1 Introduction

Taxicab service plays a vital role in public transportation by offering passengers quick personalized destination service in a semi-private and secure manner. A 2006 study reported that 241 million people rode New York City Yellow Medallion taxicabs and taxis performed approximately 470,000 trips per day, generating $1.82 billion in revenue. This accounted for 11% of total passengers, an

** This work was partially supported by the National Science Foundation under Grant No. IIS-1017926.

estimated 30% of total public transportation fares, and yielded average driver income per shift of $158 dollars [1]. Taxicab drivers earn this by cruising the road network looking for a passenger at designated taxi stands or alongside the streets. However, this service is often inefficient from expensive vehicles with low capacity utilization, high fuel costs, heavily congested traffic, and a low ratio of *live miles* (miles with a fare) to *cruising miles* (miles without a fare).

With higher fuel costs and decreasing budgets, pressure mounts on taxicab drivers who directly derive their income from fares yet spend anywhere from 35-60 percent of their time cruising the road network for fares [1]. The unpredictable nature of passengers and destinations make efficient systematic routing a challenge. Therefore, the goal is simultaneously reducing cruising miles while increasing live miles, thus increasing profitability, without systematic routing.

This paper presents a simple yet practical method for suggesting profitable locations that enable taxicab drivers to reduce cruising miles. The concept uses the same principle that a taxicab driver uses: follow your experience. We propose a framework to guide taxi drivers in locating fares. Specifically, this paper makes three contributions. First, the proposed framework uses historical GPS data to model the potential profitability of locations given the current location and time of a taxi driver. This model considers the main factors contributing to the profitability: time and the profit loss associated with reaching a location. Second, this framework makes personalized suggestions to a taxi driver based on location and time. This avoids the problem of communicating the same information to all drivers, which may result in non-equilibrium in supply and demand. Third, we demonstrate the effectiveness of the proposed framework using a large dataset of Shanghai taxicab GPS traces and use correlation to compare the suggested locations with actual driver behavior.

2 Related Work

Taxicab service falls into two general categories and research follows this, occasionally attempting to bridge them. The first category is *dispatching* where companies dispatch taxicabs to customer requested specific locations. A request may be short-term (e.g., a customer requests a taxi for pickup within the next 20 minutes) or long-term (e.g., arrangements come hours or days in advance). Logic dictates that the farther in advance the request, the easier it is to plan efficient taxi service because routing algorithms already exist (mostly based on Dijkstra's work); the shorter the request time, the more challenging the routing problem. The second category is *cruising*. The taxicab driver cruises the road network looking for a fare at designated taxi stands or alongside the streets, using experience as a guide. This leads to an inefficient system where taxi drivers spend significant time without a fare and often serve hot spots, leading to a supply and demand imbalance. Since cruising is a profit loss, this paper will refer to non-live miles as a *cruising trip* and live miles as a *live trip*. The following highlights some recent research in this area.

Yamamoto et al. propose a fuzzy clustering based dynamic routing algorithm in [2]. Using a taxicab driver's daily logs, the algorithm creates an optimal route solution based on passenger frequency on links (i.e., paths). The routes, not intended to be used directly by the taxi driver, are shared among taxis through mutual exchanges (i.e., path sharing) that assigns the most efficient path to a taxi as they cruise. This potentially reduces the competition for a potential fare, excessive supply to popular areas, and traffic congestion while increasing profitability. Similarly, Li et al. present an algorithm using taxi GPS traces to create a usage based road segment hierarchy from the frequency of taxis traversing a road segment [3]. This hierarchy inherently captures the taxi driver experience and is usable in route planning. In these two examples, the focus is on routing but trip profitability—a key factor in the driver's decision—is not explicitly addressed.

Another example of taxicab routing is T-Drive, developed by Yuan et al. to determine the fastest route to a destination at a given departure time [16]. T-Drive uses historical GPS trajectories to create a time-dependent landmark graph in which the nodes are road segments frequently traversed by taxis and a variance-entropy-based clustering approach determines the travel time distribution between two landmarks for a given period. A novel routing algorithm then uses this graph to find the fastest practical route in two stages. The first stage, *rough routing*, searches the graph for the fastest route for a sequence of landmarks; the second stage, *refined routing*, creates the real network route using the rough route. Similar to the previous example, this system does inherently capture taxi driver experience and suggests faster routes than alternative methods; however, this method does not suggest profitable locations for taxicabs.

A thesis by Han Wang proposes a methodology for combining short-term and long-term dispatching [4]. If a customer's starting and ending locations follow the path of a taxi as it heads to a different dispatch call, the taxi can pick up the fare. This allows a reduction in cruising and an increase in profitability. The catch is that it may not be common for passenger routes to align exactly. Therefore, Wang proposes the Shift Match Algorithms that suggest drivers and/or customers to adjust locations, creating a reasonable short delay in service but an improvement overall. In this study, the cruise trips are different from those in the aforementioned cruising category because they result from dispatching, not from intent to cruise. This method is practical for dispatching but not for general cruising.

Another approach, given by Cheng et al., focuses on customer queuing at taxi stands and taxis switching between serving stands and cruising [5]. Phithakkitnukoon et al. developed an inference engine with error based learning to predict vacant taxis [6] while Hong-Cheng et al. studied travel time variability on driver route choices in Shanghai taxi service [7]. Additional research covers a variety of issues from demand versus supply to pricing issues [8–12]; however, these studies do not consider location profitability, which is inherent to the driver's decision.

The work most similar to ours is by Ge et al., who provide a novel technique in extracting energy-efficient transportation patterns from taxi trajectory traces

and a mobile recommender system for taxis [17]. The technique extracts a group of successful taxi drivers and clusters their pick-up points into centroids with an assigned probability of successful pick-up. The resulting centroids become the basis for pick-up probability routes that the system distributes among taxis to improve overall business success. The major contribution is how the system evaluates candidate routes using a monotonic Potential Travel Distance (PTD) function that their novel route-recommendation algorithm exploits to prune the search space. They also provide the SkyRoute algorithm that reduces the computational costs associated with skyline routes, which dominate the candidate route set. This recommender system potentially improves success by using probabilities; however, probabilities can be misleading in relation to the profitability since high probabilities do not necessarily translate into highly profitable live trips. In addition, the algorithm clusters locations using fix periods regardless of when the taxicab actually arrives at a location. Furthermore, our framework suggests a customized map of locations based on the taxicab's current location to eliminate route creation cost and taxi-route assignment distribution issues; however, their algorithms could enhance our framework by suggesting customized paths for the taxi driver using a time series of STP maps.

3 Methodology

The taxicab driver is not concerned with finding profitable locations during a live trip. Once the live trip is complete, assuming there is not a new passenger available at the location, the driver must decide where to go. They may stay in that general vicinity for a time in hopes of a passenger or, more likely, head to another location based on experience. At this moment, the driver considers two variables: profitable locations and reasonable driving distances. However, a driver might be unaware of both variables. For example, a driver may be unreasonably far from a highly profitable airport but unsure of closer profitable locations less often visited. Given a map identifying these locations, the driver can make an informed decision quickly and reduce cruising time.

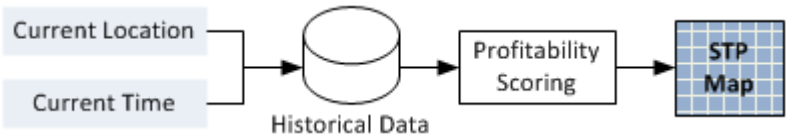


Fig. 1. The STP map generation process. The current location and time are parameters for retrieving historical data that becomes profitability scores in an STP map.

Figure 1 summarizes the proposed methodology for identifying these locations. When a taxicab begins a cruising trip, the current location and time are parameters for querying a historical database that serves as driver experiences.

The experience information coincides with locations and becomes a location-based profitability score. The process assembles these scores into an STP map that suggests potentially profitable locations to the taxicab driver. By following the suggestions, the driver can reduce cruising time thus increase profitability.

STP map generation occurs when a taxicab is ready for a new fare, i.e. when the driver begins a cruising trip. At this moment, and based on the current location, the map encompasses a region of interest within a reasonable driving distance and uses historical data to determine the profitability of locations within this region. The map is personalized to the driver since each driver is at a different location. This mechanism can prevent sending the same information to multiple drivers, which could result in localized competition and a non-equilibrium state. It is possible for multiple drivers to receive the same STP map if their closeness is within error bounds of the distance calculations, but this occurs infrequently.

This region can be large enough to encompass all the historical data, but since the taxicab moves spatially and temporally, it is not necessary to model the entire region. The first step in generating this map is to define a sub-region M around the taxicab's current location in region R such that $M \subseteq R$. In other words, region M is for short-term planning, the taxicab driver's inherent process—the driver moves towards locations of high live trip probability and profitability. Figure 2 demonstrates this concept. At time t_1, the taxi driver drops off the passenger at the end of a live trip and receives STP map $M1$ of the surrounding location. The driver chooses a profitable location within the region, moves to that location, and picks up a passenger. This new live trip continues until t_2 when the passenger is dropped off and the driver receives a new STP map. This process continues until the taxi goes out of service. With this knowledge, the driver can reduce overall cruising time.

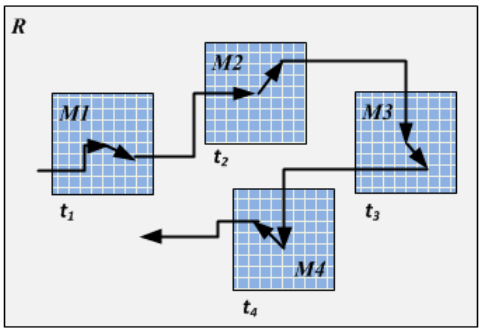


Fig. 2. A taxicab moving through region R. At each time t_i, the driver receives a new STP map M customized to that location and time with the taxicab located at the center.

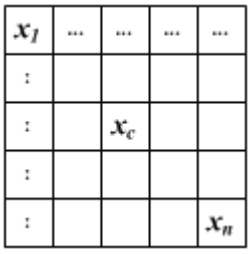


Fig. 3. An example sub-region M composed of n cells with the taxi located at x_c. Each x_i represents a location with a profitability score.

This method defines locations within M and determines a profitability score for each location. The simplest implementation is to divide M into a grid of equally sized cells such that $M = \{x_1, x_2, ..., x_n\}$, with the taxicab located at center cell x_c (see Figure 3). The grid granularity is important. The cell sizes should be large enough to represent a small immediate serviceable area and to provide enough meaningful historical information to determine potential profitability. For instance, if the cell of interest x_i has little or no associated historical information, but the cells surrounding it do, it may be beneficial to increase the granularity. On the other hand, it should also not be too large as to become meaningless and distorted in terms of profitability. For instance, a cell the size of a square kilometer may be unrepresentative of a location.

As mentioned previously, the historical data determines the cell profitability since it captures the taxi drivers' experiences in terms of trips. The natural inclination is to use the count of live trips originating from the cell as the profitability indicator; however, this can be misleading since it does not consider the probability of getting a live trip and because a trip fare calculation, which determines profitability, uses distance and time, implying that the average of some distance to time ratio for a location's trips may be more appropriate. This ratio still would not capture true profitability because trip fares are not a direct ratio of distant to time. It is common practice to charge a given rate per distant unit while the taxi is moving and a different rate when idling. Therefore, a location's profitability is a factor of trip counts (cruising and live), trip distances, and trip times (idling and moving).

Table 1. A summary of the variables used to determine profitability.

Variable	Description
d_l	Total distance of a live trip
$D(x_1, x_2)$	Distance between locations x_1 and x_2
$F(j)$	The fare of trip j
n_c	Number of cruise trips
n_l	Number of live trips
$P(x_c, j)$	Profitability of trip j
$P(x_c, x_i)$	Profitability of an STP map location
r_l	Charge rate per unit of distance
r_i	Charge rate per unit of idle time
s	Taxicab speed
t	Total time of a trip
t_c	Time of a cruise trip
t_i	Time of taxi idling
t_l	Time of a live trip
x_c	The taxicab location, which is the center of M
x_i	Location of interest within M
θ	Proportion relating unit cost
ϵ	Adjustment for low trip counts

The formula for calculating the fare for live trip j, $F(j)$, at starting location x_i is the amount of idle time $t_i(j)$ charged at rate r_i, plus the distanced traveled $d_l(j)$ charged at r_l[3] (see Eq. 1). If unknown, one can estimate idle time t_i using the total trip time t and the taxicab speed s as $t - d_l/s$. The cost of reaching the starting point of trip j is the distance traveled between x_c and x_i, $D(x_c, x_i)$, times some proportion θ of r_l since the rate charged has the cost factored into it. Therefore the profitability of trip j with the taxi currently located at x_c, $P(x_c, j)$, is $F(j)$ minus the cost associated with reaching x_i from x_c (see Eq. 2). The profitability of location x_i with respect to current location x_c, $P(x_c, x_i)$, is the sum of all historical live trip fares from that location divided by the total count of trips, both live n_l and cruising n_c, minus the cost between x_c and x_i (see Eq. 3). Eq. 1 is actually a simplification of the fare pricing structure which can vary for different cities, different locations within the city, and different times of day. The fare price may be fixed; for example, the price from JFK airport in New York is fixed, or more commonly, there is a fixed charged for a given distance, a different charge per distance unit for a bounded additional distance, and a third charge after exceeding a given distance. As an example, Table 2 gives the Shanghai taxi price structure.

$$F(j) = (t_i(j) * r_i) + (d_l(j) * r_l) \tag{1}$$

$$P(x_c, j) = F(j) - (D(x_c, x_i) * r_l * \theta) \tag{2}$$

$$P(x_c, x_i) = (\sum_{j=1}^{n_l} F(j))/(n_l + n_c) - (D(x_c, x_i) * r_l * \theta) \tag{3}$$

Table 2. The Shanghai taxicab service price structure [15]. Taxi drivers charge a flat rate of 12 Yuan for the first 3 kilometers plus a charge for each additional kilometer.

Trip Description	5am - 11pm	11pm - 5am
0 to 3 km	12.00 Total	16.00 Total
3 to 10 km	12.00 + 2.40/km	16.00 + 3.10/km
Over 10 km	12.00 + 3.60/km	16.00 + 4.70/km
Idling	2.40/5 minutes	5.10/5 minutes

It is difficult to use these equations without detailed historical information; however, there is an exploitable relationship. Time can represent the profitability by converting each variable to time; distance converts to time by dividing by the speed and the rate charged for idle time is proportional to the rate charge for movement time. Furthermore, the profit earned by a taxi driver is directly proportional to the ratio of live time t_l to cruising time t_c. The average of all

[3] Note that r_l is some proportion of r_i.

live trip times originating from x_i, minus the cost in time to get to x_i from x_c, can represent the profitability score of cell x_i (see Eq. 4). The ϵ ensures that the profitability reflects true trip probability when trip counts are low.

$$P(x_i) = (\sum_{j=1}^{n_l} t_l)/(n_l + n_c + \epsilon) - (D(x_c, x_i) * r_l * \theta)/s \quad (4)$$

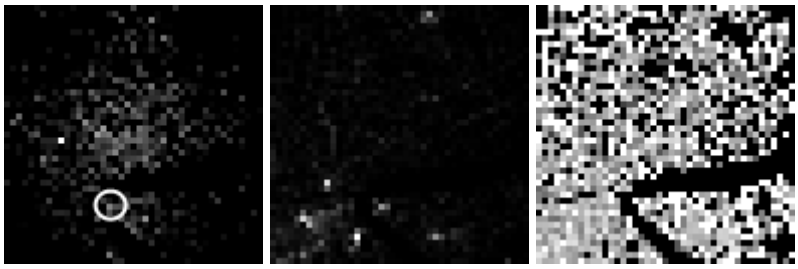
Each of these variables is derivable from GPS records. Given a set of records with an indication of the taxicab's occupancy status, the time stamps and status can determine the live and cruising times, the GPS coordinates determine the distances, and the distances and times determine speeds.

It is not a requirement to use all the historical data from the GPS records to determine the location profitability used in the STP map; in fact, using all the data may be misleading due to the changing conditions throughout the day. For example, profitability for a specific location may be significantly different during rush hour than during night traffic. Since the taxicab driver is looking for a fare in the here and now, a small data window will better represent the driver's experience for this period. For each location, the data selected should represent what the conditions will be when the driver reaches that location. For example, if it is 1:00pm and takes 10 minutes to reach the location, the historical data should begin at 1:10pm for the location. The size of this *Delayed Experience Window* (DEW) may be fixed or variable as necessary, but the size is important. If the DEW is too large, it may include data not representative of the profitability; if too small, it may not include enough data. The following case study gives an example of the proposed methodology applied to Shanghai taxicab service.

4 Case Study - Shanghai Taxi Service

Shanghai is a large metropolitan area in eastern China with over 23 million denizens [13] and a large taxicab service industry with approximately 45 thousand taxis operated by over 150 companies [14]. To demonstrate our method, we use a collection of GPS traces for May 29, 2009. The data set contains over 48.1 million GPS records (WGS84 geodetic system) for three companies between the hours of 12am and 6pm and over 468,000 predefined live trips of 17,139 taxicabs. We divided the data into the three companies and focused on the first company, which yielded data for 7,226 taxis. The region R was limited to 31.0°-31.5° N, 121.0°-122.0° E to remove extreme outliers and limit trips to the greater metropolitan area. Furthermore, only trips greater than five minutes are included since erratic behavior occurred more often in those below that threshold. Similar erratic behavior occurred with trips above three hours, often the result of the taxi going out of service, parking, and showing minute but noticeable movement from GPS satellite drift. The three-hour threshold is partially arbitrary and partially based on the distribution of trips times. While relatively rare, there are times when taxis spend over an hour on a cruising trip, but cruising trips over three hours occur much less frequently.

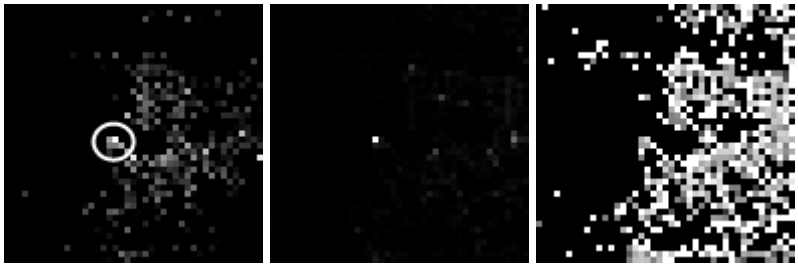
These reductions left 144 thousand live trips remaining for the first company from which we constructed cruising trips. For each taxicab, we defined the cruising trips as the time and distance between the ending of one live trip and the beginning another. We assumed there is a cruising trip before the first chronological live trip if there is at least one GPS record before the live trip starting time that indicated no passenger in the vehicle. We also assumed, that the end of the taxi's last live trip indicated that the taxicab was out of service and did not incur any additional cruising trips. For example, if the taxicab first appears at 12:04am, but its first live trip is at 12:10am, 12:04-12:10am became a cruising trip. If the taxicab's last live trip ended at 3:07pm, this became the last trip considered. This resulted in 948 fewer cruise trips than live trips, but did eliminate all outlier trips outside the period.



(a) STP Map (b) Live Trip Counts (c) Live Trip Probability

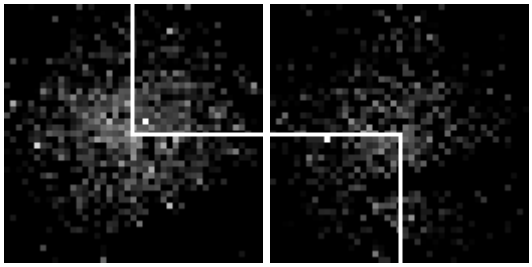
Fig. 4. Results for the downtown region at 1:00pm with a 60-minute DEW, 190.5-meter cell length, M size of 67.1 km^2, and with the taxicab in the center. The Oriental Pearl Tower is encircled. Using the live trip counts or probability could cause the taxicab driver to incur a higher cost compared to using the STP map.

For the first demonstration, the square region around the taxicab location was approximately 8.1x8.1 km^2 divided into 43x43 square cells of approximately 190.5 meters in length. The DEW is 60 minutes, starting at a time delay based on the time required for to reach the cell. We chose the taxi's current time as 1:00pm since the 1:00-3:00pm period has the largest percentage in data distribution. The time and distance required to reach the cell's center came from the L_1 Manhattan distance and average speed of 11 km/h based on instantaneous speeds recorded in the GPS data. The distances of the live trips originating from the cell is the sum of L_2 Euclidean distances from the trip's individual GPS records. For the profitability score, θ was deduced from the data to be approximately 0.333; although the exact value is unknown, it can be estimated by analyzing trip times. Figure 4 and Figure 5 display the STP maps near the downtown region and near the Shanghai Hongqiao International Airport, respectively, with the taxicab at the center of the map. The figures also include the live trip counts and probability for the areas as a comparison to the profitability. Additionally,



(a) STP Map (b) Live Trip Counts (c) Live Trip Probability

Fig. 5. Results for the Shanghai International Airport region at 1:00pm with a 60-minute DEW, 190.5-meter cell length, M size of 67.1 km^2, and with the taxicab in the center. The airport terminal is encircled. The high count of live trips from the terminal shadows the other locations, hiding other potentially profitable locations that our STP map captures.



(a) STP Map 1 (b) STP Map 2

Fig. 6. Results for overlapping downtown regions at 1:00pm with a 60-minute DEW, 90.5-meter cell length, M size of 67.1 km^2, and with the taxicab in the center. The top-right quadrant of STP Map 1 overlaps the bottom-left quadrant of STP Map 2. There is a clear difference between the overlapped regions as lower profitability areas in one are often higher profitability areas in the other.

Fig. 7. Color scale from low to high values for Figures 4, 5, and 6.

to show that two taxicabs at the same time get two distinct STP maps, Figure 6 shows two overlapping regions in which the top-right quadrant of Figure 6(a) overlaps the bottom-left quadrant of Figure 6(b). For visualization purposes, negative profitability areas are set to zero to highlight the profitable regions, which are of interest to the taxi driver.

Figure 8 overlays the STP map in Figure 4(a) with the downtown area using Google Earth and one-hour DEW. The results show a correlation with office

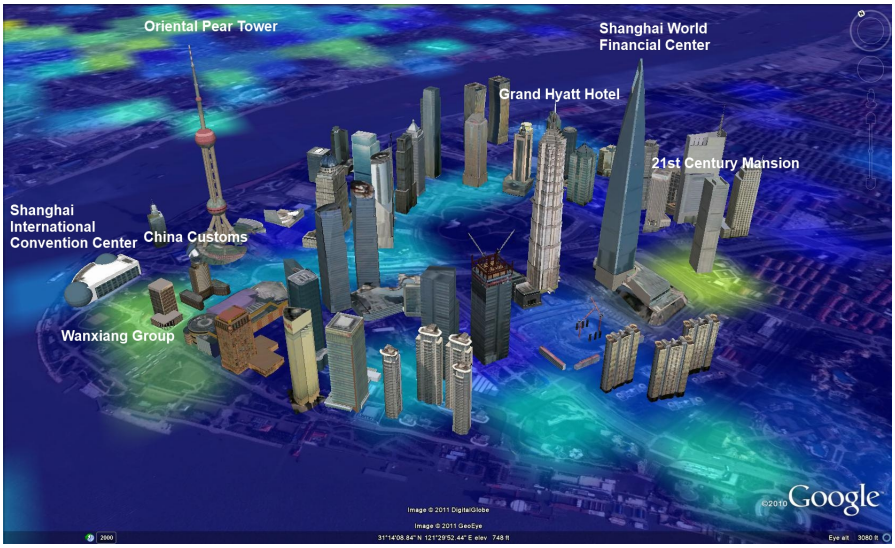


Fig. 8. Results for the downtown region STP map over laid with the Google Earth's satellite image at 1:00pm, a 60-minute DEW, 190.5 meters cell length, and M size of 67.1 km^2. Lighter areas represent higher profitability scores and often correlate with areas expected to be profitable, such as the Shanghai International Convention Center.

buildings and the STP map. There are two issues to note. First, Google Earth distorts the cell edges in an effort to stretch the image over the area, leading to potential misinterpretation, although minor. Second, the cell granularity plays an important role in the results. Near the image center, the construction area near the Grand Hyatt Hotel shows high profitability while the Grand Hyatt itself does not show as high profitability as would be expected. This is because the cell boundary between these locations is splitting the trips between them. While this is an issue, it is more typical for a group of close cells to have similar profitability scores. From the viewpoint of a taxicab driver, this is not an issue because the goal to find general locations of high profitability, not necessarily the specific 190.5 by 190.5 square meters. Figure 9 similarly shows an STP map overlaying the airport. The airport is one of the hottest locations, producing numerous profitable trips that make it a favorite location among taxi drivers. In this case, the entire terminal area has similar profitability even though the cells maybe splitting the activity among them. A graphical glitch is preventing the red cell from completely showing near the image center.

5 Validation

To validate this method, we must show that the STP maps correlate with actual profitability. If assumed that taxicab drivers move towards high profitable areas when cruising, then it is logical that the ending location of a cruising trip (i.e.,

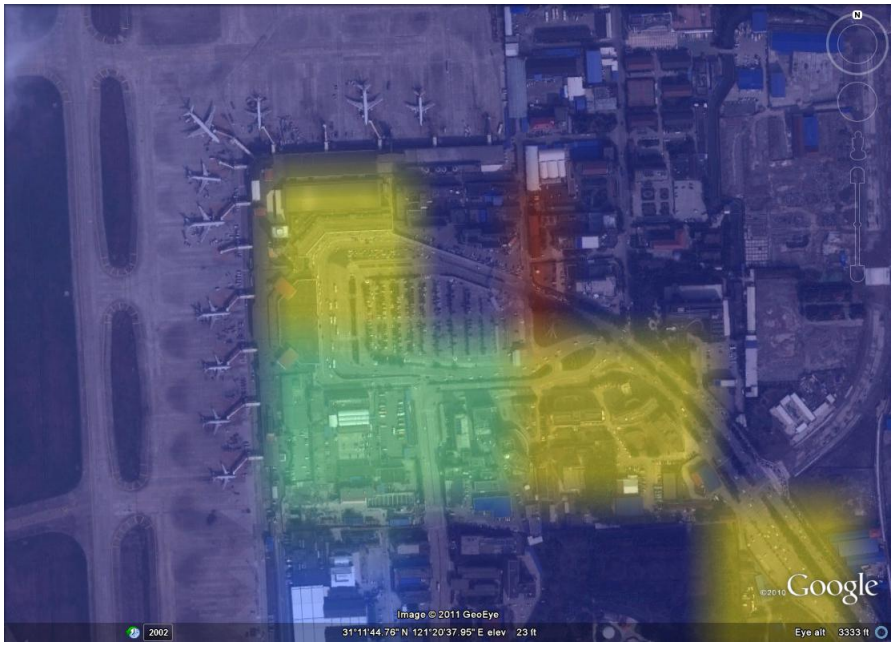


Fig. 9. Results for the Shanghai International Airport region STP map over laid with the Google Earth's satellite image at 1:00pm, a 60-minute DEW, 190.5 meters cell length, and M size of 67.1 km^2. Lighter areas represent higher profitability scores and correlates with airport terminal and surrounding area. Note that a graphical glitch is causing the red center cell to be distorted.

the beginning location of a live trip) is a profitable location. If these ending locations correlate to the higher profitable areas in the STP maps generated for the taxicab throughout the day, and this correlates with known taxi profitability, then the STP map correctly suggests good locations. In other words, if the aggregate profitability scores associated with the ending locations of cruise trips throughout the day correlates to actual profitability, which can be determined by live time to total time for a taxi, then the correlation should be positive.

We selected five distinct test sets of 600 taxicabs and removed taxis with less than 19 total trips to focus on those that covered the majority of the day. This resulted in 516-539 taxis per test set. For each taxicab, we followed their path of live and cruising trips throughout the day. When a taxicab switched from live to cruising, we generated an STP map using a 15-minute DEW for a 15.8 by 15.8 km^2 area divided into 167 by 167-square cells (approximately 95 meters in length) with the taxi at the center. We summed the profitability scores at cruising trip ending locations and correlated them with the real live time to total time ratio that defines actual profitability. We then repeated the experiment, increasing the cell size and DEW while holding the region size constant.

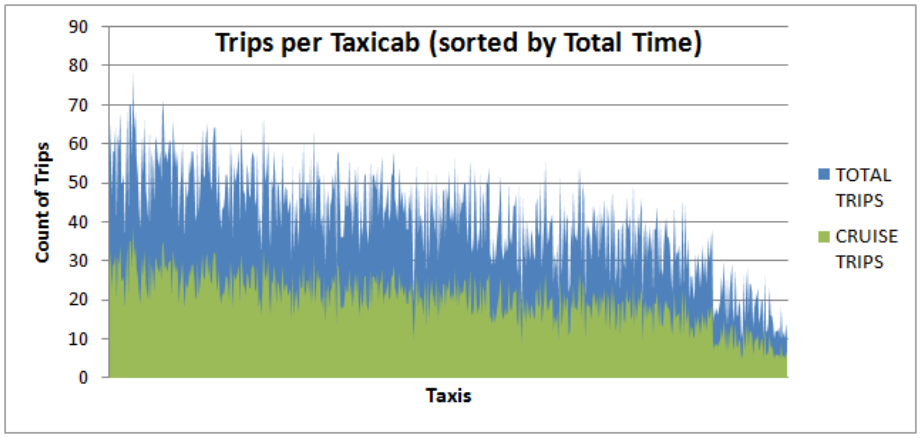


Fig. 10. The total and cruise trips for one dataset with 538 taxis, sorted by total time. Increasing the total time tends to disproportionately increase the count of cruise trips to live trips.

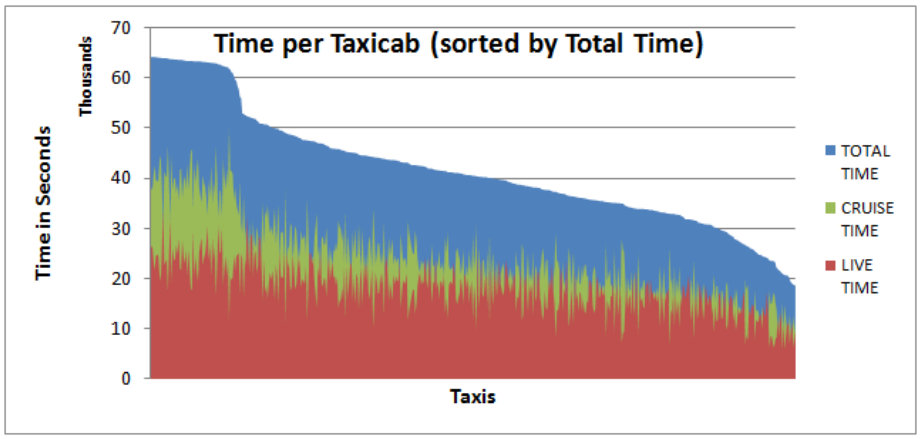


Fig. 11. The total, live, and cruise times in seconds for one dataset with 538 taxis, sorted by total time. The amount of cruising time is often greater than live time.

Figures 10, 11, and 12 visualize typical characteristics of the test sets. Figure 10 shows trip counts and Figure 11 shows trip times with taxis sorted by total time. There is a distinct group having higher total times, but this results from a larger percent of cruising time relative to the other taxis. Comparing this with Figure 12, the taxicab profits with taxis sorted by total time, reveals that the total time in service does not necessarily improve profits; in fact, it has a tendency to have the opposite effect. Figure 12 also shows that an increase in total time typically yields more trips, but does not necessarily increase overall profits.

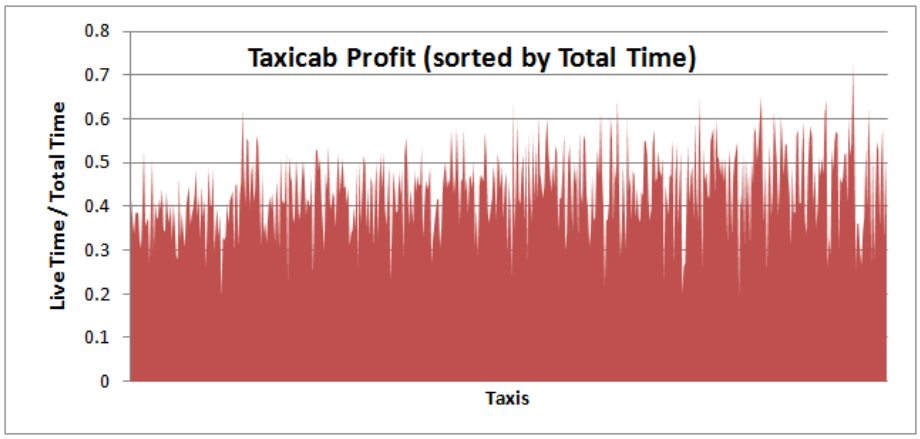


Fig. 12. The profit (live time/total time) for one dataset with 538 taxis, sorted by total time. There is a slight upward trend in profits as the total time decreases, indicating that an increase in total time does not guarantee an increase in profit.

Figure 13 displays the resulting average correlation over the datasets for three cell sizes and DEWs. The average correlations approached 0.50 with a slightly higher median. The trend in correlation clearly demonstrates the effect of cell sizes. Small sizes do not accurately represent the profitability and larger sizes tend to distort. Additionally, the DEW shows a definite trend. The more historical data, the better the correlation; however, caution should be taken. Increasing the DEW increases the amount of historical data, but may cause it to include data not representative of the current period. For example, if the DEW includes both rush hour and non-rush hour traffic, then the profitability may not reflect real profitability. In addition, if a taxi only cruises for a few minutes, the extra 50 minutes of a 60-minute DEW has less importance in making a decision. Figure 14 confirms this hypothesis—holding the 190x190 m^2 cell size constant, the correlation increases with the increasing DEW until past the 90-minute mark. Since the DEW starts at 1:00pm but was time delayed as described in the method, it started including the traffic pattern beyond the afternoon rush hour but before the evening rush hour.

The positive correlation was not as high as expected, but investigating the scatter plots revealed that there is a good correlation. As an example, Figure 15 shows the scatter plot correlation for one test set with a 30-minute DEW and a 190x190 m_2 cell size. The Hit Profit is the sum of all profit scores from cells where the taxicab ended a cruising trip and the Live Time/Total Time is the profitability of the taxicab for that day. As indicated by the trend line, the higher the taxicab profitability, the higher the Hit Profit. While the correlation for this specific set was 0.51, there is a definite upward trend in correlation among all sets with the majority of taxis are ending cruise trips in the higher profitable locations based on our method.

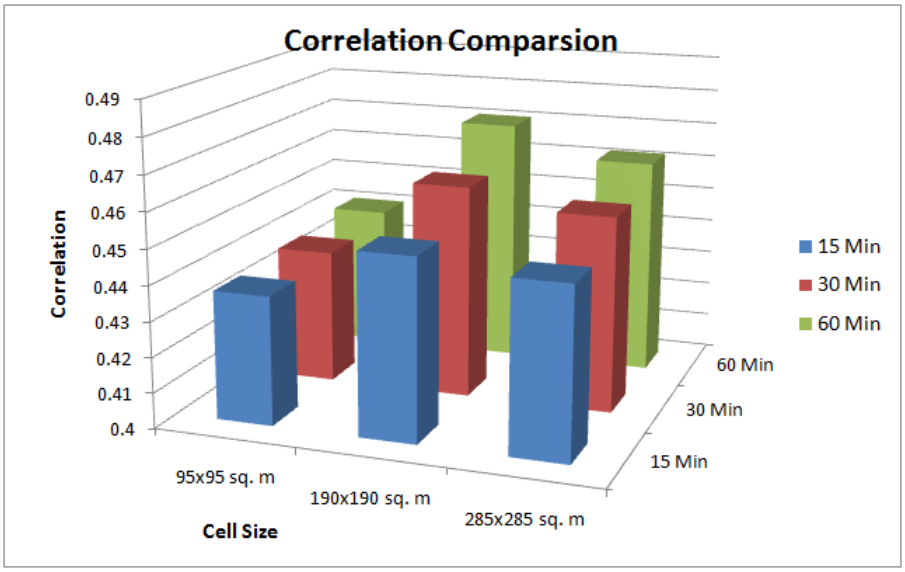


Fig. 13. Average correlation of the five test datasets for a given cell size and DEW. The cell size 190x190 m^2 produced the best overall correlation, reaching 0.51 for one of the five datasets.

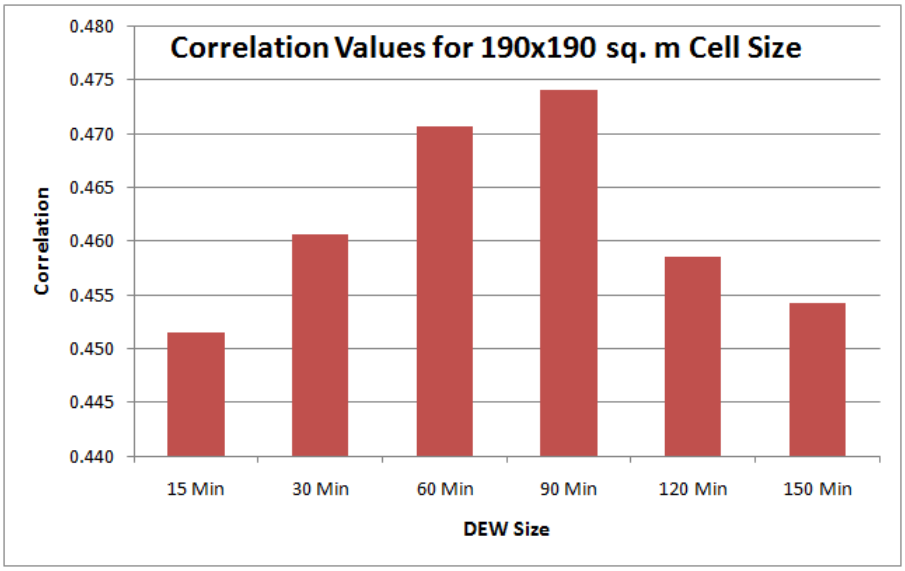


Fig. 14. Correlations values for the 190x190 m^2 cell size with varying DEW size starting at 1:00pm. A DEW size greater than 90 minutes causes a significant decrease in correlation, demonstrating it's importance.

Fig. 15. An example of correlation results with an × marker indicating a taxicab. This test set used a cell size 190x190 m^2 with a 30-minute DEW. The Hit Profit is the sum of all profit scores from cells where the taxicab ended a cruising trip and the Live Time/Total Time is the profitability of the taxicab for that day. The majority of taxicabs are ending in the more profitable locations, providing a positive correlation between the STP map and actual profitability as shown by the upward trend line.

6 Future Work

There are several potential improvements for this method. First, we did not focus on the temporal aspect beyond shifting the DEW at a delayed time and adjusting the size. Patterns in time may affect results by allowing it to include data from two distinct periods in relation to the traffic pattern; for example, rush hour traffic data included with non-rush hour data. To a lesser extent, the cell sizes and region M may need to adjust with time as well; for example, late at night, there may be a need to increase the cell size due to lower probability of live trips and an increase in M to include more potential locations. For validation purposes, M was large enough to ensure that all cruising trips considered ended within the area with our profitability scores; otherwise, the score would be zero when in reality it should be positive or negative. The goal would be to develop a dynamic STP mapping system that adjusts each of these components given current conditions and time.

Another improvement involves the distance calculations. The L_1 Manhattan distance formula determined the distance between the current taxi location and the location of interest. While this is more realistic than using the L_2 Euclidean distance, it relies on a grid city model, which is not always applicable to all

areas of the city. The live trip distances used the L_2 Euclidean distance between individual GPS records to determine total distance, which has an associated error in accuracy as well. These calculations also did not consider obstacles; for example, drivers must cross rivers at bridges or tunnels, which may add distance and time to a trip. One potential solution is to use the road network, current traffic conditions, and known obstacles to find the best path to a location and then use the path to determine profitability. An alternative is to capture the driver's intuitive nature to find the best path or to use distances of common paths traveled by multiple taxis. A preliminary investigation into this alternative revealed that it is a possibility given enough GPS records.

Since the ultimate goal is for a taxi driver to use the STP map, the system needs to be real-time and use visuals that are easy to understand and not distractive to driving. It could also take into consideration the current traffic flow to determine a more accurate profitability score, and give higher potential profitability path suggestions leading to a profitable location. This could increase the probability of picking up a passenger before reaching the suggested location and thereby further reduce cruise time and increase profits.

7 Conclusion

The growing demand for public transportation and decreasing budgets have placed emphasis on increasing taxicab profitability. Research in this area has focused on improving service through taxi routing techniques and balancing supply and demand. Realistically, cruising taxicabs do not easily lend themselves to routing because of the nature of the service and the driver's desire for short-term profitability. Since the live and cruising times define the overall profitability, and a taxicab may spend 35-60 percent of time cruising, the goal is reducing cruise time while increasing live time. Our framework potentially improves profitability by offering location suggestions to taxicab drivers, based on profitability information using historical GPS data, which can reduce overall cruising time. The method uses spatial and temporal data to generate a location suggesting STP map at the beginning of a cruise trip based on a profitability score defined by the live time to total time profitability definition. A case study of Shanghai taxi service demonstrates our method and shows the potential for increasing profits while decreasing cruise times. The correlation results between our method and actual profitability shows a promising positive correlation and potential for future work in increasing taxicab profitability.

References

1. Schaller Consulting: *The New York City Taxicab Fact Book*, Schaller Consulting, Brooklyn, NY, 2006, available at http://www.schallerconsult.com/taxi/taxifb.pdf
2. Yamamoto, K., Uesugi, K., and Watanabe, T.: *Adaptive Routing of Cruising Taxis by Mutual Exchange of Pathways*, Knowledge-Based Intelligent Information and Engineering Systems, 5178 (2008) 559–566

3. Li, Q., Zeng Z., Bisheng, Y., and Zhang, T.: *Hierarchical route planning based on taxi GPS-trajectories*, 17th International Conference on Geoinformatics (Fairfax, 2009), pp. 1–5
4. Wang, H.: *The Strategy of Utilizing Taxi Empty Cruise Time to Solve the Short Distance Trip Problem*, Masters Thesis, The University of Melbourne, 2009
5. Cheng, S. and Qu, X.: *A service choice model for optimizing taxi service delivery*, 12th International IEEE Conference on Intelligent Transportation Systems, ITSC '09 (St. Louis, 2009), pp. 1–6
6. Phithakkitnukoon, S., Veloso, M., Bento, C., Biderman, A., and Ratti, C.: *Taxi-aware map: identifying and predicting vacant taxis in the city*, Proceedings of the First international joint conference on Ambient intelligence, AmI'10 (Malaga, 2010), pp. 86–95
7. Hong-Cheng, G., Xin, Y., and Qing, W.: *Investigating the effect of travel time variability on drivers' route choice decisions in Shanghai, China*, Transportation Planning and Technology, 33 (2010) 657–669
8. Li, Y., Miller, M.A., and Cassidy, M.J., *Improving Mobility Through Enhanced Transit Services: Transit Taxi Service for Areas with Low Passenger Demand Density*, University of California, Berkeley and California Department of Transportation, 2009.
9. Cooper, J., Farrell, S., and Simpson, P.: *Identifying Demand and Optimal Location for Taxi Ranks in a Liberalized Market*, Transportation Research Board 89th Annual Meeting (2010)
10. Sirisoma, R.M.N.T., Wong, S.C., Lam, W.H.K., Wang, D., Yan, H., and Zhang, P.: *Empirical evidence for taxi customer-search model*, Transportation Research Board 88th Annual Meeting, 163 (2009) 203–210
11. Yang, H., Fung, C.S., Wong, K.I., Wong, S.C.: *Nonlinear pricing of taxi services*, Transportation Research Part A: Policy and Practice, 44 (2010) 337–348
12. Chintakayala, P., and Maitra, B., *Modeling Generalized Cost of Travel and Its Application for Improvement of Taxies in Kolkata*, Journal of Urban Planning and Development, 136 (2010) 42–49
13. Wikipedia, *Shanghai — Wikipedia, The Free Encyclopedia*, 2011, Online; accessed 21-May-2011, http://en.wikipedia.org/w/index.php?title=Shanghai&oldid=412823222
14. TravelChinaGuide.com, *Get Around Shanghai by Taxi, Shanghai Transportation*, 2011, Online; accessed 9-February-2011, http://www.travelchinaguide.com/cityguides/shanghai/transportation/taxi.htm
15. Shanghai Taxi Cab Rates and Companies, *Kuber*, 2011, Online; accessed 9-February-2011, http://kuber.appspot.com/taxi/rate
16. Yuan, Jing and Zheng, Yu and Zhang, Chengyang and Xie, Wenlei and Xie, Xing and Sun, Guangzhong and Huang, Yan, *T-drive: driving directions based on taxi trajectories*, Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '10 (San Jose, 2010), pp. 99–108
17. Ge, Yong and Xiong, Hui and Tuzhilin, Alexander and Xiao, Keli and Gruteser, Marco and Pazzani, Michael, *An energy-efficient mobile recommender system*, Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '10 (Washington, DC, 2010), pp. 899–908