Estimating Taxi Demand-Supply Level using Taxi Trajectory Data Stream

Dongxu Shao, Wei Wu, Shili Xiang, Yu Lu
Institute for Infocomm Research
A*STAR, Singapore
Email: {shaod,wwu,sxiang,luyu}@i2r.a-star.edu.sg

***Abstract*—Taxis provide a flexible and indispensable service to satisfy the urban travel demand of public commuters. Understanding taxi supply and commuter demand, especially the imbalance between the supply and the demand, would directly help to improve the quality of taxi service and eventually increase a city's traffic system efficiency. In this paper, we consider the taxi demand from a region during a period of time to include two parts: satisfied demand, i.e., passengers successfully receive taxi service during this period of time, and unmet demand, i.e., passengers are still waiting for taxi service. To properly estimate the demand-supply level (short for "the level of the taxi demand vs. supply imbalance"), we propose a novel indicator that reflects how fast an available taxi is taken in any given region. Accordingly, we design and implement a taxi analytics system to provide such information in near real time. Finally, we use the passenger waiting time survey data and the taxi streaming data to validate the proposed indicator on the built taxi analytics system.**

I. INTRODUCTION

Taxi, providing personalized and efficient point-to-point service, is a popular form of public transport in many cities. Especially, in the densely populated Asian cities, taxi service is often at a low cost as well, and thus many people rely on taxis for their daily trips, to fulfil their business, social, and family activities. The taxi service quality not only affects the level of satisfaction of locals to the land transportation services, but also has an impact on the experience of visitors and tourists to the convenience of the transport facilities.

While there are a large number of taxis on the road, it happens that passengers have difficulties in finding available taxis in certain areas and hours. Fortunately, with advances in technology in sensor and wireless communications, nowadays taxis are commonly equipped with GPS (Global Positioning System) receivers and wireless communication components (such as GSM, GPRS). With these devices, taxis easily reports the location and status (such as FREE, HIRED) information to the central server, for the taxi operators to explore to improve their quality of service, by real-time location tracking and call dispatching to meet the needs of time-sensitive users, etc. To make taxi drivers and commuters more aware of the real-time demand and supply, Land Transport Authority (LTA) of Singapore has also recently launched a mobile app called Taxi-Taxi@SG [11], which displays the locations of available taxis to commuters in real time and also enables commuters to broadcast their current locations to drivers.

How to match the supply with the demand is of critical importance, for the interest of both passengers and taxi drivers, both taxi operators and land transport authorities. There are many existing research works in this direction, such as modelling the spatio-temporal structure of taxi services [3], [5], predicting the demand at taxi stands [10], or even recommending the next-passenger to the drivers [6]. There are also literatures focusing on passengers' perspective, such as to predict where to find empty taxis [12]. However, all the above works are understanding the number of demand as the number of historical pickups, while in realities the historical pickups are constrained by the historical supplies. There could be unmet demand (i.e., unsatisfied passenger demand still waiting for a taxi service) not captured by the pickup records. Moreover, the above works are either suggesting taxi's passenger-look-up strategy based on status of other taxis and historical pickups, or suggesting passengers based on location of free taxis, without focusing on the relationships of the two parties. Therefore, in this paper, we will endeavour to infer the demand and supply imbalance, by bearing in mind the demand as both successful pickups and unmet demand, to make a more complete picture for the passengers and taxis. According to our knowledge, this is the first piece of work that deals with the demand-supply relationship in regions.

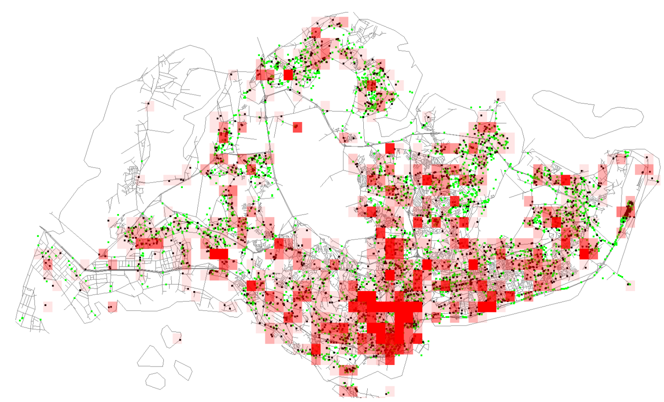


Fig. 1. A Snapshot on the Distribution of Taxis in Singapore

In this paper, we adopt a data-driven approach to model the taxi demand and supply imbalance. We design a taxi data analytics system that efficiently processes the taxi trajectory and status data, intelligently analyses it, derives insights such as demand-supply level, and displays the insights in real time. Let us briefly illustrate the idea of our system with Figure 1. Figure 1 shows the snapshot of the distribution of Taxis at a particular time in Singapore. Each green dot represents an available taxi, while each black dot represents a taxi with passenger on board. It shows that some areas have plenty of available taxis with only a few taxis taken, while the situation is opposite in some other areas. This suggests a demand-supply imbalance in various regions around Singapore. Our system intend to estimate the demand-supply level in each region(represented by each square), and display the demand-supply level through intensity of some color (such as red in the figure), to help drivers and passengers to make their decisions.

We propose an indicator to estimate the demand-supply level, based on how fast FREE taxis are taken. This is motivated by the idea that the free taxi should statistically have higher probability to meet a demand sooner if the demand is high while the supply is low in the region. Figure 2 shows the trajectory of a taxi that enters and exits a region and its status change (green denotes FREE while red denotes HIRED), and our classification of demand-supply level, to illustrate the above idea of our proposed approach. Basically, we study the behaviour of each taxi in the region in micro level. Each pickup is not considered as a discrete event, but as a result of a sequence of cruising. Instead of treating the number of pickups as a whole, we focus on the status transition for the pickups and the time taken for the transition, and aggregate such information for all the taxis inside that area to make a good sense of the whole region's demand-supply level.

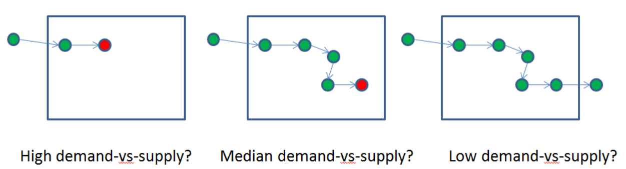


Fig. 2. Demand-Supply Illustration

We summarize the main contribution of this paper as follows:

- We identify an important problem of modelling demand-supply imbalance, where the demand refers to both successful pickups and the unmet passenger demand.
- We propose a novel indicator to estimate the demand-supply level, by measuring how fast FREE taxis are taken. We also design a methodology to quantify the indicator with the observations from taxi live data, which maximizes the log-likelihood function.
- We design a taxi live data analytics system to implement our proposed methodology in real time, validate it against the surveyed average passenger waiting time (PWT) data, and the experimental results demonstrate that the probability is indeed a proper indicator for the demand-supply level.

The rest of the paper is organized as follows. Section 2 introduces our datasets, which are used to derive insights or conduct validation. Then, in Section 3, we depict our overall taxi live data analytics system architecture, and describe how the data extraction works efficiently to facilitate demand-supply level estimation in real time. Our methodology to estimate the demand-supply level is described in detail in Section 4, and validated in Section 5. Afterwards, Section 6 present some related works on taxi trajectory and status data analysis. At last, we conclude with future works in Section 7.

II. Taxi Trajectory and Status Data

A. Taxi Data

Each taxi in Singapore is equipped with a specific in-vehicle telematics device, called mobile data terminal (MDT). MDT can automatically collect taxi's real time GPS location, taxi status and other critical information. The current MDT device is able to explicitly identify multiple taxi states, and the taxi data we collected including six taxi states as below:

- FREE - Taxi unoccupied and ready for taking new passengers or bookings;
- BUSY - Taxi driver temporarily unavailable due to a personal reason;
- HIRED - Passenger on board and taximeter running;
- ONCALL - Taxi unoccupied, but accepted a new booking job;
- CHANGE SHIFT - Taxi Driver is on the way to change shift;
- OTHERS - Other feasible taxi states.

All the real time taxi information, including the current taxi state and GPS location, with the corresponding time stamp are automatically sent to the backend system via the general packet radio service (GPRS) module on MDT. The MDT update frequency is typically 30 seconds.

B. Passenger Wait Time Data

To collect feedback from Singapore taxi commuters, the Land Transport Authority (LTA) conducts monitoring and checks by collecting passenger related information at hourly intervals between 5pm-12am at major taxi stands across the island. At approximately 40 taxi stands, LTA conducts the passenger waiting time survey: the surveyors manually record down each passengers arrival time and boarding time at the taxi stands. If several passengers leave with one taxi, they would be regarded as a group and the group size is also marked. Meanwhile, the surveyors also take down all the arrival taxis state (e.g., TAXI (refering to FREE in the MDT state, or ON CALL (refering to ONCALL in the MDT state), and if a waiting passenger finally selects to leave the taxi stand for other transportation options, such event would be also recorded. In this paper, we use the survey data from January 2015 to March 2015 to validate our findings.

III. System Overview

The work presented in this paper is part of a big taxi live data analytics system prototype where a number of aspects of the taxi system are analyzed including demand-supply level, taxi usage, impact of weather, and driver strategy etc. Figure 3 shows the high level overview of system components that are relevant to taxi demand-supply level estimation and validation.

The MDT device in each taxi sends a record to a server every t seconds (see previous section for detail of data record).

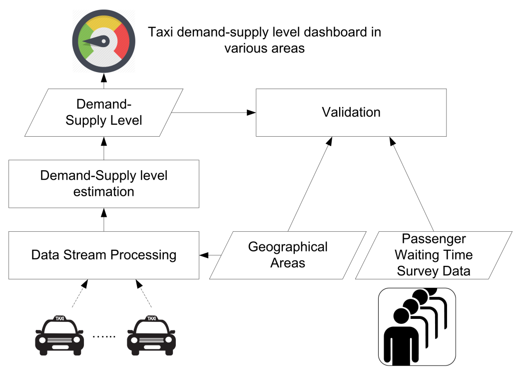


Fig. 3. System Overview

The geographical areas of interest are modelled as polygons and they are an input of the system. The data stream processing component takes data streams from taxis and extract key events in each geographical area that will be useful in understanding taxi demand-supply level. The demand-supply level estimation component takes the key events to estimate the real time demand-supply level for each geographical area. The resultant demand-supply level in each area is visualized on a map to form a taxi demand supply information dashboard.

As mentioned in the previous section, the Land Transport Authority (LTA) of Singapore sends surveyors to selected taxi stands to collect passenger waiting time information. Passenger waiting time can be a proxy of the taxi demand and supply level. The manually collected PWT data provide valuable ground truth of taxi demand-supply level for us to validate the result produced by our system using taxi data streams. As a result, our system prototype also contains a validation component that checks the demand-supply level against PWT data. In the validation process, geographical areas are also taken as input. The taxi stands (where PWT data are available) in each area are used to compute the average waiting time in the area.

For each area, the data stream processing component extracts the following information for each taxi that is observed in FREE status inside the area: ID, T1, T2, Status (or Exit). ID is the unique identifier of the taxi, T1 is the first time stamp when the taxi is observed in FREE status inside the area. T2 is the time stamp when the taxi is non-free (i.e., become HIRED, or BUSY, etc) inside the area, or the time stamp when the taxi exits the area in FREE status (i.e., the taxi exits the area without getting a passenger).

Figure 4 shows a state and event diagram that describes how the data stream processing component extracts key information from taxi live data stream to facilitate demand-supply level estimation. When a record is received from a taxi, the system checks whether the taxi has entered in a new geographical area. Once the taxi enters a geographical area (under monitoring), the system checks whether the taxi is in FREE status or not. If the taxi is in FREE status, the time the taxi enters the area is recorded as T1, otherwise the status of the taxi will be continuously monitored. If the taxi changes to FREE inside the area, that time stamp is recorded as T1. Basically, the first time stamp the taxi is observed in FREE status inside the area is recorded as T1. Then the location

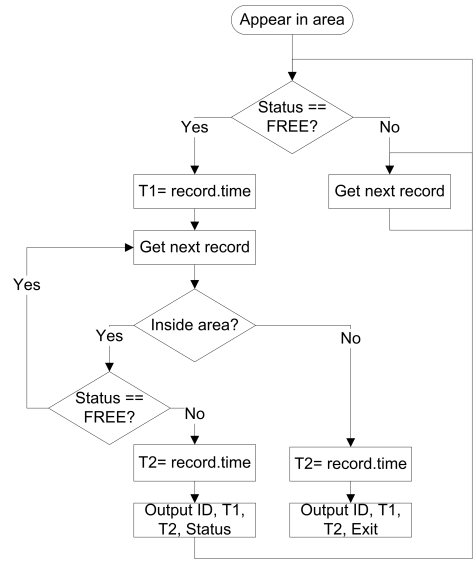


Fig. 4. State Diagram

and status of the taxi are monitored to check whether the taxi becomes non-free or whether the taxi exits the area. When such event happens, the time stamp of that event is recorded as T2 and the type of event is also recorded. In more detail, if a taxi becomes HIRED at T2, the output event record will be (ID,T1,T2,HIRED); if it becomes BUSY at T2, the output event record will be (ID,T1,T2,BUSY); if it exits the region as FREE at T2, the output event record will be (ID,T1,T2,Exit). Each taxi will be monitored until it exits the geographical area under monitoring, and it is possible that multiple event records will be extracted and output in the same geographical area.

The extracted information tells when a taxi supply is available inside a area and when that taxi supply is taken (if the status becomes HIRED) or not taken (if it exits as FREE) by a demand, or it is temporarily not available inside the area (if its status is neither HIRED nor FREE). If the supply is taken, (T2-T1) also tells us how much time it took the taxi to meet a demand. We will not be able to estimate demand-supply level via a single taxi's data, but by aggregating such information from a large number of taxis we hope to make a good sense of the demand-supply level inside the area.

Note that our system is able to monitor multiple areas while processing data from a large number of taxis. In our prototype for Singapore, our system processes live data from more than 28,000 taxis and the Singapore island is divided into hundreds of geographical areas and our program monitors the demand-supply level for each of these areas. The geographical areas are indexed in an in-memory R-Tree to speed up spatial operations (e.g., checking whether a taxi is inside an area).

IV. METHODOLOGY

The ultimate goal is to estimate the demand-supply level of a given area in a time window. While the number of supply can be obtained by monitoring the taxis with status FREE, the number of demand cannot be observed. So we have to derive the demand level from the observations of taxi live data.

Let us imagine the scenario of an area with a number of demand. A FREE taxi appearing in this area will drive along the road until it meets a demand. If the number of demand is relatively high, then the density of the demand on the road is high. So the FREE taxi will meet a demand sooner. Motivated by this, we propose a novel model to estimate the demand-supply level by making some assumptions.

Assumption 1: All the demands are generated in the beginning of each time window, and uniformly distributed in the area.

Assumption 2: The taxis are driving with identical speed within the area and time window.

Let p be the probability that a FREE taxi meets a demand within one time unit. The above assumptions guarantee this parameter is well-defined. This indicator can be interpreted as how fast a FREE taxi is taken in one time unit.

Suppose the duration of one time unit is T minutes. Hereafter when talking a timestamp t, we mean the $t-$th time unit from the beginning. Therefore the free duration of a taxi refers to the number of time units when the taxi is FREE.

Consider the observation that a FREE taxi appearing at time t_1 and becoming HIRED at time t_2. Such observation can be extracted as a pair $(x, 1)$, where the first element $x = t_2 - t_1$ is the duration of FREE status and the second element indicates the taxi is HIRED at the end of its FREE period. This observation can be decomposed to a series of events that the taxi does not meet a demand in time unit $t_1, \ldots, t_2 - 1$, and meets a demand at time t_2. Hence the probability that this observation happens conditioning on p can be calculated as follows.

$$Pr[(x, 1)|p] = (1 - p)^x \cdot p. \tag{1}$$

Generally a FREE taxi may end within the time window with different types.

1) The taxi becomes HIRED.
2) The status of the taxi becomes other than FREE and HIRED, such as BUSY.
3) The taxi leaves this area within the time window with FREE status.

Definition 1: Given a FREE taxi, its free pair (x, y) is the pair of number of free time unit x, and ending type y, where $y = 1$ if the taxi ends with HIRED and $y = 0$ otherwise.

Hence the probability of a free pair (x, y) conditioning on p is

$$Pr[(x, y)|p] = (1 - p)^x \cdot p^y. \tag{2}$$

Suppose a FREE taxi appears at time t_1 and changes its status to BUSY at time t_2. Then the free duration is $x = t_2 - t_1$ and the ending type is $y = 0$. From the observation we cannot tell whether the taxi meets a demand at time t_2. So the probability of this observation is $(1-p)^{t_2-t_1}$. Suppose the taxi does not change its status but leaves the area with FREE status at time t_2, then the free duration is $x = t_2 - t_1 + 1$ since it is still FREE at time t_2 and we do not know what happens after that. So the probability of this observation is $(1 - p)^{t_2-t_1+1}$.

Now suppose we have extracted all the free pairs from the taxi live data: $D = \langle(x_1, y_1), \ldots, (x_n, y_n)\rangle$. Then the likelihood function of p based on D is

$$L(p|D) = \prod_{i=1}^{n}(1 - p)^{x_i} \cdot p^{y_i}. \tag{3}$$

So the log-likelihood function is

$$\log L(p|D) = (\sum_{i=1}^{n} x_i) \log(1 - p) + (\sum_{i=1}^{n} y_i) \log p. \tag{4}$$

The maximum of the likelihood function can be obtained by letting the derivative of the log-likelihood function be zero:

$$\frac{d \log L(p|D)}{dp} = \frac{-\sum_{i=1}^{n} x_i}{1 - p} + \frac{\sum_{i=1}^{n} y_i}{p} = 0. \tag{5}$$

Hence

$$p = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} x_i}. \tag{6}$$

We imagine two extreme cases to see how the parameter p indicate the demand-supply level. Suppose there is no demand in the area within the time window. Then no FREE taxi will end with HIRED. In other words, for any free pair (x_i, y_i), it must be the case that $y_i = 0$. Hence $p = 0$. Now suppose the number of demands is so large that every FREE taxi becomes HIRED whenever it appears. Then for any free pair (x_i, y_i), it must be the case that $x_i = 0$. Hence $p = 1$.

V. Results

In this section we measure the correlation between the parameter p and the actual average waiting time for each time window. The average waiting time can be obtained from the passenger waiting time (PWT) survey dataset.

The PWT dataset only records passenger activity between 17:00 and 23:59. So the passengers joining the queue before 23:59 and boarding after 23:59 are not included. Accordingly, we will exclude p for the last time window of the day.

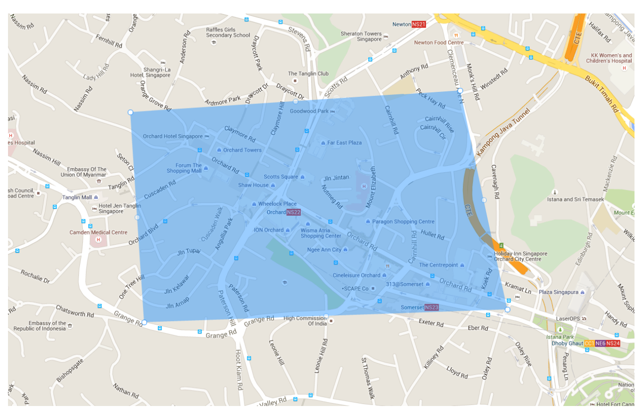
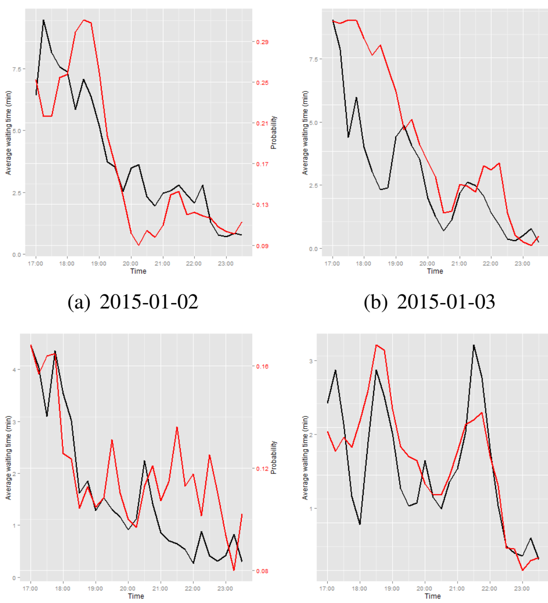


Fig. 5. Orchard Area

The PWT dataset is a survey of taxi stands, while p is designed for an area. So they are not naturally correlated. In order to correlate p with PWT data to verify our model, we select several taxi stands to form an area. The selected taxi stands have to be close to each other, and the passengers queuing at these taxi stands should be the vast majority of the demand in this area. Due to these requirements, we choose the taxi stands around Orchard road as shown in Fig. 5, where passengers are allowed to take taxis at taxi stands only.

In the experiments we set the time window length to be 15 mins. So there are 27 time windows from 17:00 to 23:45 everyday. The curves of average waiting time and probability are illustrated in Fig. 6, Fig. 7 and Fig. 8 for each day. The red curves show the probability and the black curves show the average waiting time.



(a) 2015-01-02 (b) 2015-01-03

(c) 2015-01-04 (d) 2015-01-05

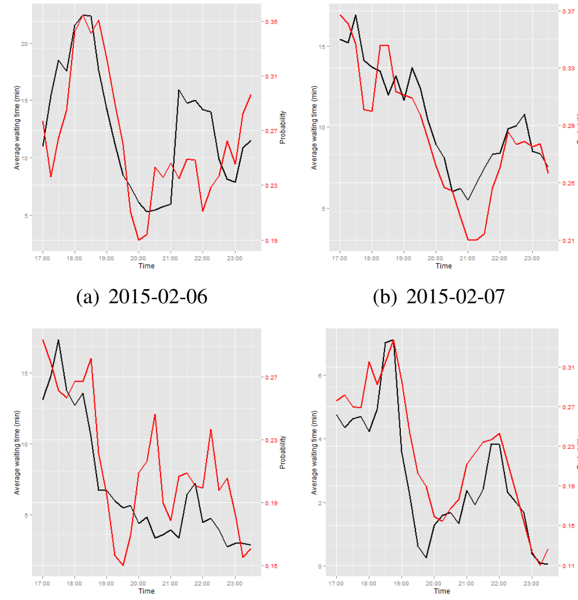
Fig. 6. The Probability and The Average Waiting Time in Each Time Window, 2015 Jan

As shown in these figures, the probability and the average waiting time have similar temporal patterns. For dates such as 2015-01-03 (Fig. 6(b)) and 2015-02-09 (Fig. 7(d)), the curve of probability and the curve of average waiting time coincidence very well, captured by the high correlations.

The correlations are summarized in Fig. 9. From the result we can see that the correlation is good (above 0.75) for most of the dates. The mean value of the correlations is 0.7760956, and the standard deviation is 0.09528578. The worst correlation happens on 2015-03-07, but we can still see the similarity from Fig. 8(b) regarding proper time wrapping.

VI. Related Works

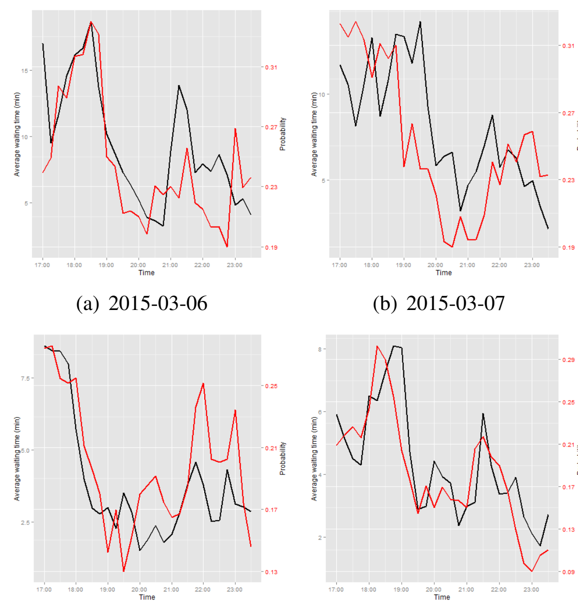
In recent years, there are a lot of research works focusing on analysing the taxi GPS trajectory and status data, to exploit its rich spatio-temporal information. The major aim is to understand and improve the performance of the taxi system in the city, and to help taxi drivers and the passengers. Several existing works analysed the taxi demand in regions, such as modelling the spatio-temporal structure of taxi services [3], [5], recommending next-passenger pick-up locations to drivers so that the drivers' profit can be maximized [6], [13]. Our work belongs to this category, but we differentiate from the existing works by modelling demand-supply relationship and considering both met demand (historical pickups) and unmet demand (yet unsatisfied passenger demand) instead of simply treating historical pickups as the demand. There are also some other works studying the taxi service at specific locations, e.g., predicting the demand at taxi stands [10], inferring the



(a) 2015-02-06 (b) 2015-02-07

(c) 2015-02-08 (d) 2015-02-09

Fig. 7. The Probability and The Average Waiting Time in Each Time Window, 2015 Feb



(a) 2015-03-06 (b) 2015-03-07

(c) 2015-03-08 (d) 2015-03-09

Fig. 8. The Probability and The Average Waiting Time in Each Time Window, 2015 Mar

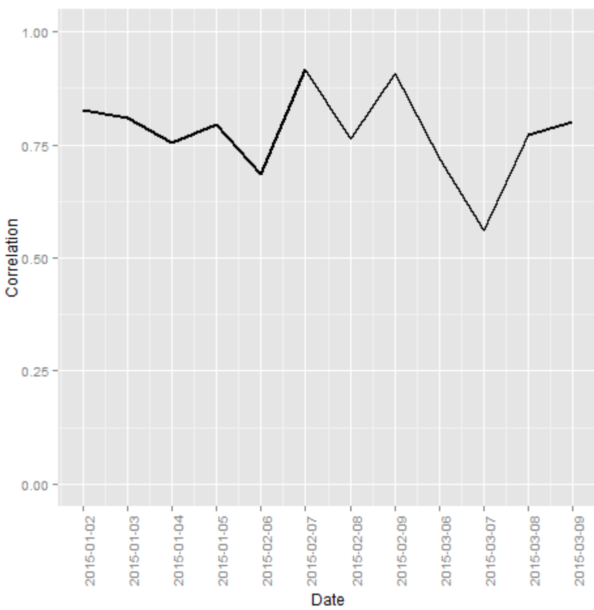


Fig. 9. The Correlation Between Probability and Average Waiting Time

relationship of supply and demand in each taxi queue hotspots [8], and allocating taxis among the few terminals in the airport [1]. In addition, there are literatures focusing more on the interest of passengers, such as to predict where to find empty taxis [12], to design recommender system to help passengers who are looking for taxis and available taxis find each other with higher probability [19]. Furthermore, recently some researchers even explored the possibilities for passengers sharing the same taxi when the demand is large in [14] and [9].

The taxi trajectory and status data is also widely analysed by multiple works and applications to study various urban problems, towards a smarter city. For example, towards building a more intelligent transport system, taxi data has been well exploited to estimate traffic speed [17], traffic volume [2], and travel time [16], and to even recommend travel routes for other drivers by discovering the intelligence of experienced taxi drivers [18], [20]. In the attempt of understanding the city dynamics, taxi data serves as a good representative as well. Through origin and destinations of the trips, taxi data tells us the mobility patterns, and attractiveness of places [7], [15]. Taxi Trajectory has also been used in conjunction with other data sources, such as mobile phone moment data, to gain better insights in population dynamics, transportation and urban configuration in [4].

VII. Conclusion and Future Work

In this paper, we design and implement a taxi data analytics system, and propose a novel indicator to help estimate the level of taxi demand-supply imbalance. Using the survey data on passengers' average wait time, we conduct the evaluation and the results show a high correlation between the indicator and the average passenger wait time.

The proposed indicator and the built system lay a solid foundation for future works, such as investigating other indicators to further refine the estimation, building a predictive model to quantify the imbalance of taxi demand and supply. It would also help to develop a real-time recommendation system for taxi drivers and passengers, to alleviate the demand and supply imbalance. Lastly, we believe that the proposed system and methodology are not only applicable to Singapore, but also for many other cities that face the similar highly dynamic imbalance of taxi demand and supply.

References

[1] A. Anwar, M. Volkov, and D. Rus, "Changinow: A mobile application for efficient taxi allocation at airports," in *16th International IEEE Conference on Intelligent Transportation Systems(ITSC)*, Oct 2013, pp. 694–701.

[2] J. Aslam, S. Lim, X. Pan, and D. Rus, "City-scale traffic estimation from a roving sensor network," in *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems(SenSys)*, New York, NY, USA, 2012, pp. 141–154.

[3] H. Chang, Y. Tai, and J. Y. Hsu, "Context-aware taxi demand hotspots prediction," *International Journal of Business Intelligence and Data Mining*, vol. 5, no. 1, pp. 3–18, Dec. 2010.

[4] C. Kang, S. Sobolevsky, Y. Liu, and C. Ratti, "Exploring human movements in singapore: A comparative analysis based on mobile phone and taxicab usages," in *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing(UrbComp)*, 2013, pp. 1:1–1:8.

[5] J. Lee, I. Shin, and G.-L. Park, "Analysis of the passenger pick-up pattern for taxi location recommendation," in *Fourth International Conference on Networked Computing and Advanced Information Management (NCM.)*, vol. 1, Sept 2008, pp. 199–204.

[6] B. Li, D. Zhang, L. Sun, C. Chen, S. Li, G. Qi, and Q. Yang, "Hunting or waiting? discovering passenger-finding strategies from a large-scale real-world taxi dataset," in *IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, March 2011, pp. 63–68.

[7] X. Li, Z. Wu, E. Qi, S. Li, D. Zhang, and W. Zhang, "Prediction of urban human mobility using large-scale taxi traces and its applications," *Frontiers of Computer Science in China*, vol. 6, no. 1, pp. 111–121, 2012.

[8] Y. Lu, S. Xiang, and W. Wu, "Taxi queue, passenger queue or no queue? - A queue detection and analysis system using taxi state transition," in *Proceedings of the 18th International Conference on Extending Database Technology (EDBT), Brussels, Belgium*, 2015, pp. 593–604.

[9] S. Ma, Y. Zheng, and O. Wolfson, "Real-time city-scale taxi ridesharing," *IEEE Transaction on Knowledge Discovery and Data Engineering*, July 2015.

[10] L. Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas, "Predicting taxi-passenger demand using streaming data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1393–1402, Sept 2013.

[11] L. T. A. of Singapore. (2014, Dec) Taxi-taxi@sg smart phone application. [Online]. Available: http://www.lta.gov.sg/apps/news/page.aspx?id=a4e14d79-21c8-49b5-b561-4326fe09af02

[12] S. Phithakkitnukoon, M. Veloso, C. Bento, A. Biderman, and C. Ratti, "Taxi-aware map: Identifying and predicting vacant taxis in the city," in *Proceedings of the First International Joint Conference on Ambient Intelligence*, ser. AmI'10, 2010, pp. 86–95.

[13] 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)*, 2014, pp. 45–54.

[14] D. O. Santos and E. C. Xavier, "Dynamic taxi and ridesharing: A framework and heuristics for the optimization problem," in *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence(IJCAI)*. AAAI Press, 2013, pp. 2885–2891.

[15] M. Veloso, S. Phithakkitnukoon, and C. Bento, "Urban mobility study using taxi traces," in *Proceedings of the 2011 International Workshop on Trajectory Data Mining and Analysis*. New York, NY, USA: ACM, 2011, pp. 23–30.

[16] Y. Wang, Y. Zheng, and Y. Xue, "Travel time estimation of a path using sparse trajectories," in *KDD 2014*. ACM, August 2014.

[17] L. Yu, J. C. P. Yee, W. S. Ng, and S. Xiang, "Data fusion using stationary detection and roaming vehicles based on speed simulation," in *The 13th ITS Asia Pacific Forum*, New Zealand, 2014.

[18] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "T-drive: Enhancing driving directions with taxi drivers' intelligence," *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, January 2012.

[19] N. J. Yuan, Y. Zheng, L. Zhang, and X. Xie, "T-finder: A recommender system for finding passengers and vacant taxis," *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, September 2012.

[20] J. Zheng and L. M. Ni, "Modeling heterogeneous routing decisions in trajectories for driving experience learning," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing(UbiComp)*, 2014, pp. 951–961.