
How does taxi driver behavior impact their profit? Discerning the real driving from large scale GPS traces

Thananut Phiboonbanakit

School of Information Computer
and Communication Technology,
Sirindhorn International Institute
of Technology, Thammasat
University, Tiwanon Road,
Pathum Thani, Thailand
m5822040068@studentmail.siiit.tu
.ac.th

Teerayut Horanont

School of Information Computer
and Communication Technology,
Sirindhorn International Institute
of Technology, Thammasat
University, Tiwanon Road,
Pathum Thani, Thailand
teerayut@siiit.tu.ac.th

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
Copyright is held by the owner/author(s).
UbiComp/ISWC'16 Adjunct, September 12-16, 2016, Heidelberg, Germany
ACM 978-1-4503-4462-3/16/09.
<http://dx.doi.org/10.1145/2968219.2968417>

Abstract

With a trend towards the use of large scale vehicle probe data, the entire urban scale analysis is become possible in order to suggest useful information for taxi drivers and passengers. This study, first, we calculate cost using cost-distance algorithm by obtained fare rate and reconstructed trips. Then, the data were analyzed to understand distance profit and service area in timely basis. The result indicated that the pickup rate of taxi in this area is usually peak according to department store operation hours. The highest chance to get customers is in the weekday rush hours between 7am and 10am and at the midnight for the weekend. The increased profits were mainly based on the distance. Finally, the hotspot of taxicabs destinations can be observed at the interchange of major public transportation. These results uncover taxi driving behavior on one of the densest network in Bangkok and yields great benefit for both taxi drivers and passengers.

Author Keywords

Taxi GPS; travel behavior; traveling profit; taxi trajectory; data mining

ACM Classification Keywords

H.2.8 Database Application: data mining, spatial databases and GIS.

Introduction

Nowadays, many innovative of mobility devices have been introducing to the society and, eventually, cause many changes in our daily life. People could use their device to serve their needs, and interact with them as a personal assistant. From this aspect, we could use the mobile technology to recognize people even better from their mobile data and trajectory movements. Global positioning system or GPS is widely used in this way. It is a standard component of today smartphone for detecting the individual current location and, even, speed of travel. It can be used to determine individual trajectory pattern and routing selection. The taxicab plays an important role in the transportation network of the city. It has been determined as an important basic transportation in most capital city around the world. However, in many countries, the economic issues as well as the presence of alternative modes of transportation cause a negative impact on taxi driver incomes.

Many researchers tried to find the way to solve this issue. Hwang et al [1], propose a taxi recommended system for determining the next cruising location by using L-L graph model and their findings suggested that distance and waiting time were more important factors than an average revenue or transition probability to determine the routing suggestion. Qu et al [2], developed a cost-effective recommended system for taxi drivers. With this method, the profits seemed to be increased if the drivers followed the recommended routes to find the passengers. The profits also included the energy efficient per driving hours. Also, the recursive strategy can help to efficiently identify the recommended optimal routes. Kamimura et al [3], present a recommendation system, called D-Taxi,

which would inform taxi drivers where to find the next passenger using the latest picking-up/dropping-off and positional information from other taxis. The aim of the researchers was to reduce the vacancy time of taxis, generate more profit for the drivers and reduce the waiting time of passengers. Ding et al [4], define a new method called global-optimal trajectory retrieving (GOTR). The study indicated a connected trajectory of high profit and high probability to pick up a passenger within a given time period in real-time. Zhang et al [5], proposed a novel method of pick-up recommendation for taxi driver based on spatio-temporal clustering. The GPS points were clustered at the different time and different regions created by taxi-drivers based on spatio-temporal clustering (RPTBSC). Salanova et al [6], presented a review of the different models developed for the taxicab problems. The models were divided into two categories, namely, aggregated and equilibrium models. Each model was analyzed from different points of view, such as market organization, operational organization and regulation issues. Zheng et al [7], mines interesting locations and classical travel sequences in a given geospatial region and GPS trajectory by using tree-based hierarchical graph (TBHG). The interesting locations mean the culturally important places, such as Tiananmen Square in Beijing, and frequented public areas, like shopping malls and restaurants. A link was generated to connect two users (a node) if they had visited more than 5 time of the same locations. The relatively big nodes denoted the top ten experienced users in Beijing. Yuan et al [8], presented a recommendation for taxi drivers and people expecting to take a taxi, using the knowledge of i) passenger's mobility patterns and ii) taxi driver's pick-up behaviors learned from the GPS trajectories of taxicabs. Qi et al [9], also presented a method to

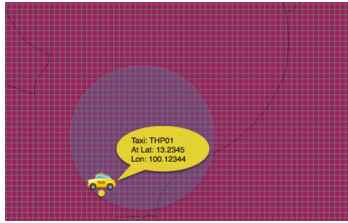


Figure 2: the taxicab is located in grid (grid detection method) to categorize location by area. The grid is 1 by 1 kilometer in size.

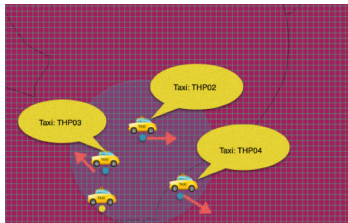


Figure 3: The historical data of previous taxicab which move around the current area.

predict the waiting time for a passenger at a given time and spot from historical taxi trajectories. They solved the problem through estimating passenger arrivals with pick-ups in the sequence of events in each spot. The concept event sequence would depict the sequential events of taxi arrival and departure in each spot with Nonhomogeneous Poisson process (NHPP). Yue et al [10], used taxi trajectory data to discover attractive areas where people often visit, for instance, hot shopping and leisure places or living and working areas based on their LoA (level of attractiveness). Data represented the frequency and density of passenger pick-up and drop-off points. People's interest to these areas varies through time of the day, day of the week, even season of the year. Zhang et al [11], propose a cruising system, pCruise, for taxicab drivers to maximize their profits by finding the optimal route to pick up a passenger. The cruising graph performed a depth-first traversal.

In Bangkok, Thailand, taxicab has played an important role in public transportation. However, many taxi drivers in Thailand have got any degree of financial problems, mostly, brought about by multifactor such as higher cost of living and energy costs, regression of Thai economic growth as well as the new incoming competitors of alternative transportation. This issue has been widely discussed by many sectors but no real solution has been reached.

The objective of the present research is to use large scale taxi probes data to discover the earning potential of taxi driver base on spatial and temporal profiles. The suggestion could provide information related to taxi behavior even better than in the past.

Methodology

The overall framework for taxi GPS data recognition and analysis was shown in Figure 1. According to the framework, it comprised of three main steps, namely, GPS Data Collection, Taxi Fare-Rate Calculation Algorithm, and Profit and Probability analysis.

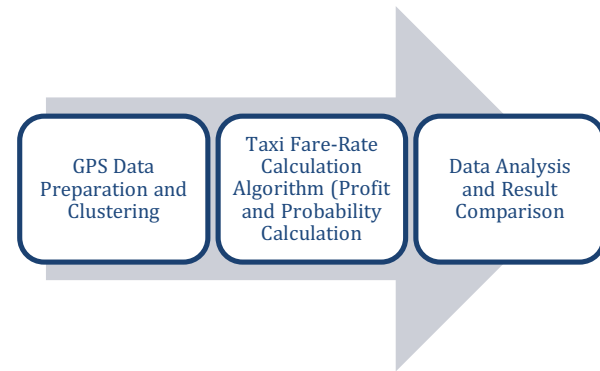


Figure 1: Research Framework

In the first Step, GPS data were collected from two thousand eight hundred and eighty two taxi probes which routed from the most famous shopping spot in central Bangkok. The data included unique ID (IMEI), location, speed, direction, date, meter status, engine, and data sources (Table 1). We developed an algorithm to calculate the taxi cost (taxi fare rate) of each individual trip of thousands of taxi for 5 months. In this step, the entire GPS points are processed and result in a format of taxi cost, distance, traveling time and date-time. It is finally grouped as trip (Origin and destination). The grid with size 1 by 1 kilometer is applied to cluster the taxi GPS points into area (Figure 2). We also analyzed historical data for further predication (Figure 3).

Field	Visualization	Notes
IMEI	10011304	Taxi Identification
Latitude	13.73522	Degree
Longitude	100.58979	Degree
Speed	30	Each move speed (km/h)
Direction	16	Degree
Error	0	Floating Point
Acceleration	0	0 represent No accelerate and 1 represent that the accelerate occur
Meter	1	0 represent free route or no passenger and 1 represent that taxi with passenger
Date-Time	2016-01-14 10:39:40	Time of each movement

Table 1: Dataset of Taxi GPS which collect from real driving vehicles in Bangkok.

An algorithm for calculating taxi cost was developed by Phiboonbanakit et al [12] based on the fare rate obtained from the Department of Land Transport (DLT). [13] The algorithm was adapted to improve the efficiency by adding traffic delay to the fare rate calculation part. The result indicates the cost, traffic delay, waiting time, trip time and timestamp of each trip. We separate trips into two categories; trip that has passenger (meter = 1) and trip of cruising to find customer (meter = 0). In the final step, the profit and probabilistic analysis were performed. The net-profit was calculated by extracting the service cost out of current profit as shown in the equation (1):

$$NetProfit(N) = Cost - \left(\frac{d+ct}{md}\right) - (T * S) \quad (1)$$

where d = distance by hour and T = total trip time
In the present study, Toyota Corolla Altis 1.6 CNG [14], which commonly used as the taxi vehicle, was used as a sample vehicle which has 55 liters of fuel tank and 75 liters of NGV tank. The fuel consumption is 12.19 km/L. The price is 22.04 Baht per liter for fuel and 13.36 Baht per liter for NGV (applied on May 5,2016 [15]). "ct" stands for cost as full tank. The calculation was shown in the equation (2):

$$ct = fueltank * fuelprice \quad (2)$$

md denoted as maximum distance which vehicle can drive from one full tank of fuel or NGV which can be calculated from the fuel consumption and total amount of fuel in one tank [14]; equation (3)

$$md = fuelconsumption * fueltank \quad (3)$$

S denoted as service cost which calculated from taxi rental cost in Thailand which is about 1,000 Baht per day [16] and divided by 24 Hours to find cost per hour and finally divided by 60 to find cost per minutes as shown in the equation (4):

$$S = \frac{\frac{taxi\ rental\ cost}{24}}{60} \quad (4)$$

To calculate Net profit per distance, we substitute "N" from equation (1) and "d" from the total distance by hour show in equation (5):

$$C = \frac{N}{d} \quad (5)$$

Then the average variable got from each trip obtained from equation (5) would be multiplied with total distance of each hour as shown in the equation (6) to get the profit:

$$Profit = \left(\frac{C1+C2+C3...+Cn}{Cn}\right) * Total\ distance\ in\ each\ hour \quad (6)$$

After that, we calculate the probability of taxi to get customer from Rama I area in each grid and group by hour. The formula to calculate the probability which taxicab will have customer onboard routing from Rama I area was shown in the equation (7).

$$p = \frac{\Sigma(\text{meter}=1)}{\Sigma(\text{all taxicap in that grid})} \quad (7)$$

Finally, the taxi destination density would be studied according to the time slot (divide into 8 hours-time slot) which was shown in table 2.

Time Slot	Name of each Time Slot
0.00 – 3.00	After Midnight
3.00 – 6.00	Early Morning
6.00 – 9.00	Morning
9.00 – 12.00	Morning with peak hour
12.00 – 15.00	Afternoon
15.00 – 18.00	Late Afternoon with peak hour
18.00 – 21.00	Evening
21.00 – 23.59	Late Evening

Table 2: Time slot for visualize taxi destination density route from Rama I and surrounding area.

Result and Discussion

From the proposed objective, we explain the result by dividing them into 4 sections.

1. Number of taxi route from Rama I area

As illustrated in Figure 4, it is shown that the number of taxis route from Rama I with passengers dropped from midnight until 5.00 am. and increased again during the day. The number of taxi peak between 11.00 am. – 02.00 pm. where many shopping stores starting to open. There was a sharp decrease in the numbers of

taxi during evening rush hours in this area which can be explained as there is a higher demand of taxi in the business area. It is therefore depending on the day of week. From the graph, the number of usage of taxi in the weekend is much higher than in the weekday. This trend is illustrating clearly in the Rama I area, as it is a major entertainment and shopping district.

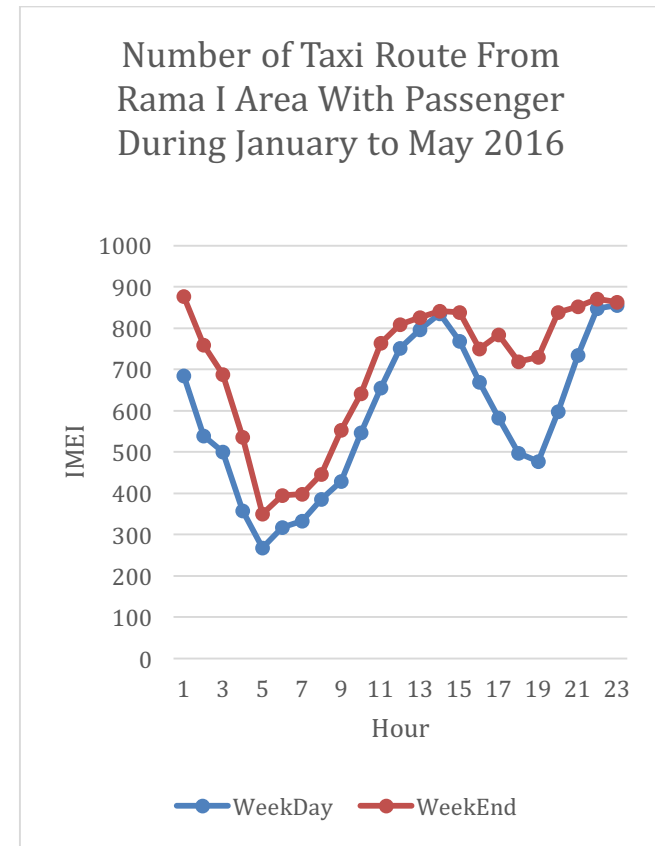


Figure 4: Number of trip route from Rama I area.

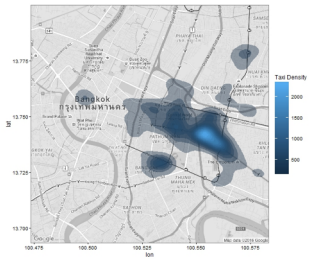


Figure 7: Destination Density of Taxicab between 0.00-3.00 h.



Figure 8: Destination Density of Taxicab between 3.00-6.00 h.

2. Net Profit per distance by hour of the day

In Figure 5, it is shown that profits would be increased directly relate to an increase in the distance. The longer distance which taxi driver obtains the higher profit they can earn. From this graph, it is shown that profit will drop from 02.00-05.00 h in the morning and rise up again from peak hours till 14.00 and another peak in the evening. Therefore, the number of taxi drop between 15.00 – 19.00 in the weekday since this area is a shopping district but if we change to the other area may change due to area of interest.

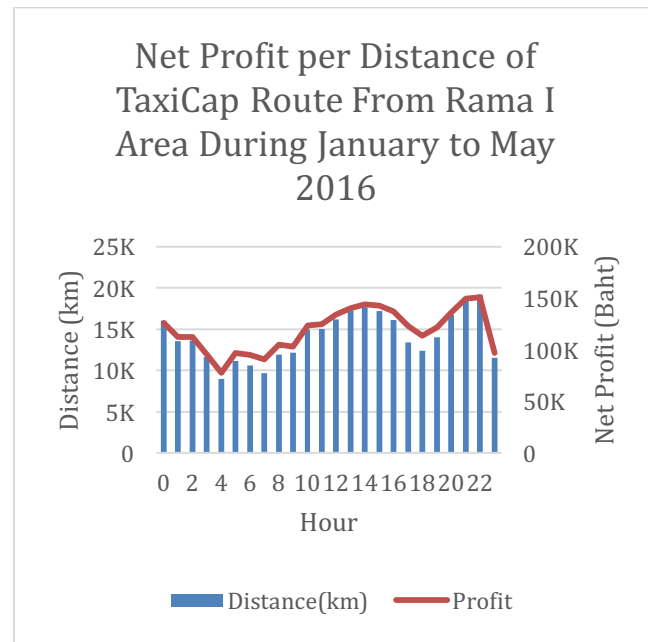


Figure 5: Cost per distance when route from Rama I area.

3. Chance to get Customer by hour of the day

Figure 6 illustrated the chance to get customer from Rama I area increases by 66% during 7.00 am. – 9.00 am. which is the week day peak hour where people are supposed to use a lot of public transportation. The probability in the weekend also increases (by 63%) in the early morning more than in the weekday, especially midnight, 1.00, 2.00, 5.00 and 6.00 am. This might be due to the facts that it is the time when people would spend their time in the weekend travel for their night event. In other times of the day, it does not have different much between weekday and weekend. Please note again that this data illustrates only in Rama I area, so if we perform an analysis at different area, the trend could be vary.

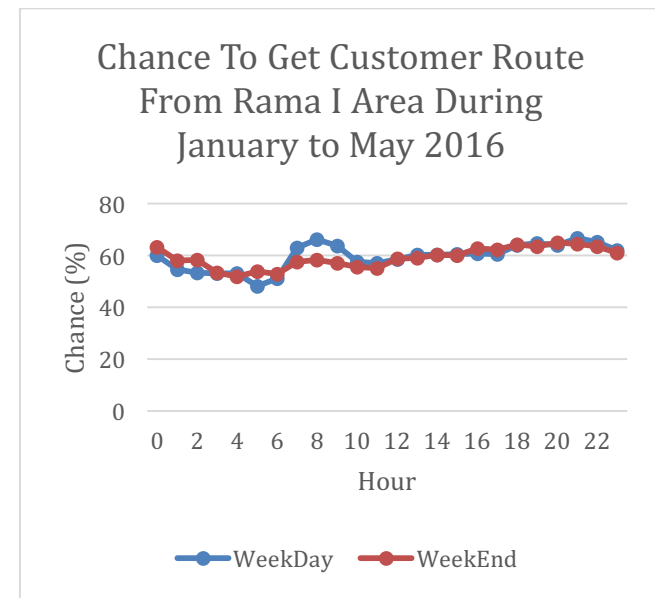


Figure 6: Chance to get customer when route from Rama I area.

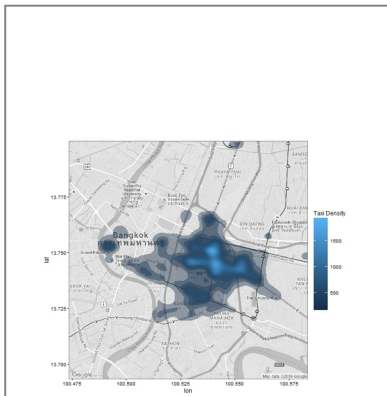


Figure 10: Destination density between 9.00 – 12.00 h. This period also contains peak hour.

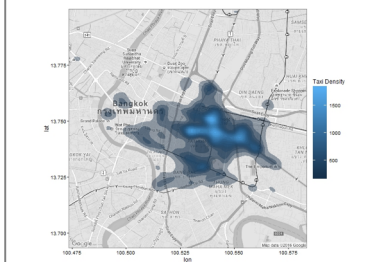


Figure 11: Destination density between 12.00 – 15.00 h.

4. Taxi destination density by time slot

In the previous section, we have described number of taxi in each hour, profit per distance by hour and the chance for taxi to get customer from Rama I area. In the last section, the density map of destinations where people routed from Rama I area to outside (Figure 7). We have divided the time into 8-time slot as mentioned earlier (Table 2). In Figure 7, after midnight time slot, the highest density area is around Nana Area (Night entertainment venue location in Bangkok). In the early morning time, the highest destination density is still at the same area but becomes widely spreading to the Ratchaprarop Road, Pechaburi Road, National BTS Sky Train Station and, finally, Bangkok main train station (Figure 8). Once look at the morning time which also cover the peak hour period (Figure 9), the density has extended to cover all of Rama I road near by shopping area. Also, an increase density nearby National Stadium and Ratchathewi BTS sky trains station, where people use sky train as transportation for traveling to working area, is observed. However, at peak hour (Figure 10), the high density area narrows down from the previous time slot. The new destination density occurs at Siam BTS Station as this time, the nearby Siam center, an important economy area, starts to open.

In the afternoon (Figure 11), most density coverage area of destination remains the same as described above. However, the density at nightlight district is decreased. Instead, the density seems to increase at National Stadium and Siam BTS, Bangkok sky train Station area. Also as shown in Figure 12 (late afternoon), the density at Siam, Bangkok shopping district and Ratchathewi BTS sky train station still remain the same but other area are reduced

dramatically. However, the density increases again around public transportation area in the evening (Figure 13). This can be explained by the fact of the peak hours when people travel back home. In the late evening (Figure 14), the destination density is still high along the Rama I area when compared to other locations. As mentioned, many shops and restaurants situated in this area and very popular among people who like to enjoy their lives after the office time.

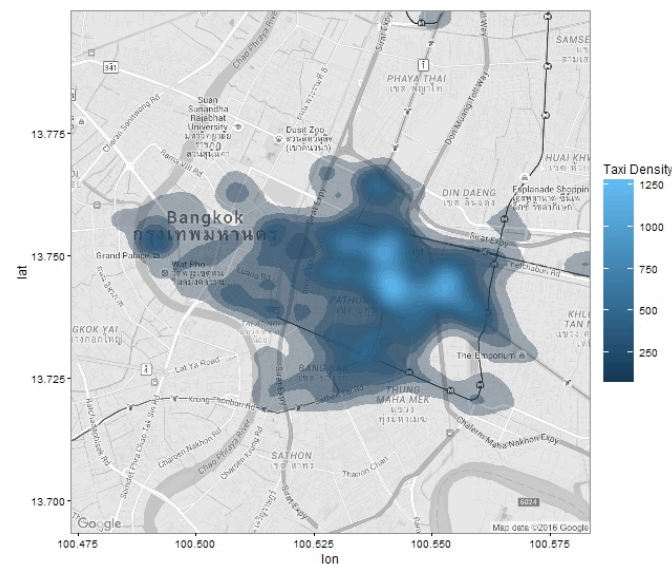


Figure 9: Destination Density of Taxicab between 6.00 - 9.00 h. The coverage area covers most area around Rama I and this period contain peak hour.

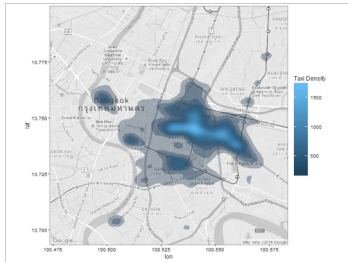


Figure 13: Destination Density of Taxicab between 18.00-21.00 h.

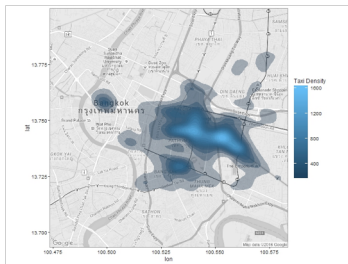


Figure 14: Destination Density of Taxicab during 21.00-23.59 h. This is the last time slot of the day. But taxicab still has routing to destination in Rama I area.

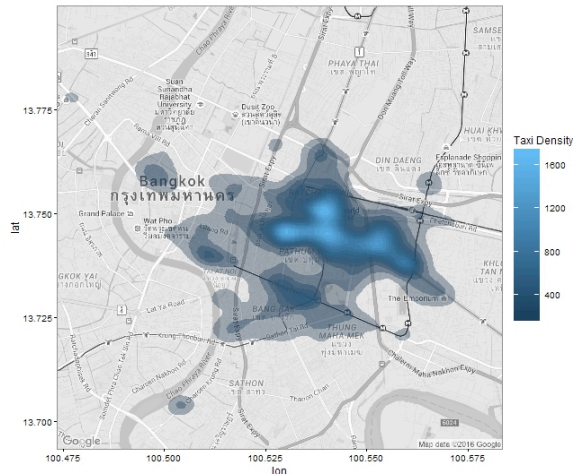


Figure 12: Destination Density of Taxicabs between 15.00-18.00 h.

Conclusion

By analyzing the Large Scale Taxi probe data, the present study comes up with some recommendations for the taxi driver to get an opportunity to earn more profits. From our findings, firstly, we recognized that, in Rama I road area, the number of taxi are increased due to the operation hour of the shopping department store since in this area is a shopping district. Secondly, profit of each taxi driver was dependent on the distance that they drive. Thus, start from 8.00 am. until 02.00 pm. and 6.00 pm. until 10.00 pm. is the suitable time where taxi driver would gain more profits than in other period. Thirdly, chance to get customer might be increased by 66% during 7.00 – 9.00 h period which is the weekday peak hour where people use a lot of transportation at that time. Also, the chance in the weekend increases (by 63%) in the early morning, especially midnight, 1.00, 2.00, 5.00 and 6.00 am. It is emphasized that

these time periods give the taxi the higher opportunity to get customers for the next route on the weekday and weekend respectively. Finally, the destination density will be high along the public transportation station, such as sky train and economic area. We found that taxicabs are acted like a transit mode connecting with other transportation. From the data, it illustrates that route to the area where has a lot of interchange of transportation is very valuable. Taxi driver can gain opportunity to increase profit in these area This result is illustrate only in Rama I area which is the shopping district. It might be change depend on the area of interest.

Acknowledgements

GPS of every dataset that we obtained are support by Toyota Tsusho Electronic (Thailand) Co, Ltd.

References

1. Ren-Hung Hwang, Yu-Ling Hsueh, and Yu-Ting Chen. 2015. An effective taxi recommender system based on a spatio-temporal factor analysis model. *Information Sciences* 314: 28-40.
2. Meng Qu, Hengshu Zhu, Junming Liu, and Guannan Liu. 2014. A Cost-Effective Recommender System for Taxi Drivers. In *Proceedings of Hui Xiong KDD'14*, 45-54. <http://dx.doi.org/10.1145/2623330.2623668>.
3. Junpei Kamimura, Hisaya Wakayama, Naoki Shiota, Masatsugu Ogawa, Norihisa Iga, and Masafumi Yano. 2013. D-Taxi: Adaptive Area Recommendation System for Taxis by using DiRAC. In *Proceedings of International Conference on Connected Vehicles and Expo (ICCVE)*, 507-508.
4. Ye Ding, Siyuan Liu, Jiansu Pu, Lionel M. Ni. 2013. HUNTS: A Trajectory Recommendation System for Effective and Efficient Hunting of Taxi Passengers.

- In *Proceedings of IEEE 14th International Conference on Mobile Data Management*, 107-116. DOI 10.1109/MDM.2013.21.
5. Mingyue Zhang, Jianxun Liu, Yizhi Liu, Zhenyang Hu, and Liang Yi. 2012. Recommending Pick-up Points for Taxi-drivers based on Spatio-temporal Clustering. In *Proceedings of IEEE Second International Conference on Cloud and Green Computing*, 67-72. DOI 10.1109/CGC.2012.34.
 6. Josep Maria Salanova, Miquel Estrada, Georgia Aifadopoulou, and Evangelos Mitsakis. 2011. *Procedia Social and Behavioral Sciences* 20: 150-161.
 7. Yu Zheng, Lizhu Zhang, Xing Xie, and Wei-Ying Ma. 2009. Mining Interesting Locations and Travel Sequences from GPS Trajectories. In *Proceedings of WWW 2009*.
 8. Jing Yuan, Yu Zheng, Liuhan Zhang, Xing Xie, and Guangzhong Sun. 2011. Where to Find My Next Passenger? In *Proceedings of UbiComp'11*, 109-118. ACM 978-1-4503-0630-0/11/09...\$10.00.
 9. Guande Qi, Gang Pan, Shijian Li, Zhaohui Wu, Daqing Zhang, Lin Sun, and Laurence Tianruo Yang. 2013. How Long a Passenger Waits for a Vacant Taxi? —Large-scale Taxi Trace Mining for Smart Cities. In *Proceedings of IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, 1019-1036. DOI 10.1109/GreenCom-iThings-CPSCom.2013.175.
 10. Yang Yue, Yan Zhuang, Qingquan Li, and Qingzhou Mao. 2009. Mining Time-dependent Attractive Areas and Movement Patterns from Taxi Trajectory Data.
 11. Desheng Zhang and Tian He. 2012. pCruise: Reducing Cruising Miles for Taxicab Networks. In *Proceedings of IEEE 33rd Real-Time Systems Symposium*, 85-94. DOI 10.1109/RTSS.2012.61.
 12. Thananut Phiboonanakit and Teerayut Horanont. 2016. Who will get benefit from the new taxi fare rate? Discerning the real driving from Taxi GPS Data. In *Proceedings of 7th International Conference on Information and Communication Technology for Embedded System*, 73-78. DOI 10.1109/ICTEmSys.2016.7467125
 13. Bangkokpost.com. Taxi fares go up in Bangkok Saturday. Retrieved August 10, 2015 from <http://www.bangkokpost.com/learning/learning-from-news/449490/taxi-fares-go-up-in-bangkok-today>
 14. Car.go.th. Car Eco Sticker. Retrieved June 9, 2016 from <http://www.car.go.th/new/compareCar?TranId=e77d299df68646ceac0407aadcdae2b7>
 15. Pttplc.com. Bangkok Oil Price. Retrieved May 11, 2016 from <http://www.pttplc.com/TH/Media-Center/Oil-Price/pages/Bangkok-Oil-Price.aspx>
 16. Thaitaxiclick.com. Taxi Car Webboard. Retrieved May 11, 2016 from <http://www.thaitaxiclick.com/?board=2.0>.