The analysis of Indian taxi behavior using long-term probe data

Riko SAKATA[1], Ashutosh KUMAR [2], Hiroshi KANASUGI[3] and Yoshihide SEKIMOTO[4]
1 Institute of Industrial Science, The University of Tokyo, 4-6-1, Komaba, Meguro-ku, Tokyo, 153-8505, JAPAN
1 School of Engineering, The University of Tokyo,7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8654, Japan
2 Department of Civil Engineering, IIT Kanpur, Kanpur 208016, India
3 Earth Observation Data Integration and Fusion Research Initiative, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8654, Japan
4 Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan

KEY WORDS: probe data, taxi behavior, big data

ABSTRACT: Nowadays chronic traffic congestion is a severe problem in Indian cities. Although some countermeasures, such as highway construction and the expansion of public transportation, were done, it doesn't seem to be enough to manage the rapid increase in traffic demand. In order to take more useful measures, grasping the traffic situation precisely seems to be needed. Looking at the metropolitan area in India, taxi is one of popular transportation. It is said that the taxi fare in Delhi is the cheapest in the world, and the middle class people often use **a** taxi in daily life. Hence, understanding the taxi behavior is important. For this purpose, probe data or floating car data is widely used nowadays. Drivers' behavior, such as their usual routes and the strategy of their business, were analyzed. In more micro-scale, some researches revealed the features which affected drivers' route choice behavior. However, these researches were done mostly in developed countries and there are few analyses in developing countries like India because of lack of data volume, although traffic situation there is totally different from that in developed countries. Therefore, the aim of this research is to get good knowledge about Indian taxi drivers' behavior using long-term probe data. For example, for what purpose do people in Indian cities use taxi, where do taxi drivers wait for the customer, and which factors would impact their such behavior. We got the taxi probe data in the whole of India for more than one-year duration and analyzed focusing on 10 metropolitan cities.

1 INTORODUCTION

The population in India have rapidly increased in recent years. According to the census, the total population in India reached 1229 million in 2016, and it seems to still increase in next several decades (Ministry of Internal Affairs and Communications Bureau of Statistics in Japan, 2017). It led to the increase in the number of vehicles shown in Figure 1 (Government of National Capital Territory of Delhi, 2011), and metropolitan cities are suffering from severe traffic jam. For example, the Ministry of Urban Development, Government of India, reported in 2007 that the average traveling speed in Delhi was about 16 km/h and only slightly higher in Mumbai (M Absar A. and Faisal A., 2013). To solve or mitigate this problem, some public institutions are executing the road expansion and implementing public transportation as countermeasures, they couldn't manage the rapid increase of traffic demand at this moment (P.S.Bhargavi & N. Kannaiya Raja, 2011).

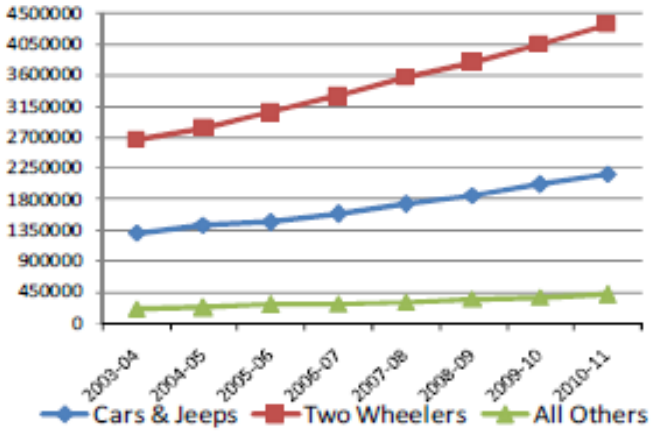

Figure 1. Vehicular Growth in Delhi(Category-wise)

In order to take more effective countermeasures against traffic congestion, revealing people's traffic behavior and accordingly thinking countermeasures might be needed. Looking at transportation, taxi fare in Indian metropolitan

cities is much cheaper than any other major cities in the world (TripAdvisor, 2011), so that we can say that taxi is one of the very popular transportation mode for the middle class people. Thus, it seems to be very important to know taxi's behavior for the solution of traffic congestion.

In recent researches, probe data is often used for the analysis of taxi or other road traffic behavior. However, such kinds of analysis as taxi drivers' business behavior, waiting spots, the strategy of getting customers are done mainly in developed countries. Yoshii et al. (2005) analyzed the taxi behavior when they are vacant based on probe data. They found that taxi drivers tend to wait for customers in taxi pool more than looking for customers while driving. Kingetsu and Hattori (2015) also studied using probe data individual taxi drivers' strategy for getting customers. They found that the strategy would be different from each driver and built a model expressing the strategy. On the other hand, in developing countries, such as India, these kinds of researches seem to be very few because of lack of the environment for data acquisition or some reasons. Some research has examined countermeasures against congestion in Hyderabad (Aparajita C, Sudakshina G, 2014). The study discussed the pros and cons of possible countermeasures based on literature review and statistical report, but the criteria are based on just quality and quantitative approach seems not to have been fully examined. Other research aims to get insight about the mitigation of traffic jam in Indian cities, but they analyzed just statistical data and they didn't mention the difference of impact of each transportation (Bhuvanachithra C, Dimitrios Z, 2012). We can imagine that developed countries and developing countries have different traffic situation, in terms of road capacity, establishment level of public transportations, road registrations like driver's license. Thus, analyzing the traffic behavior on major transportation in developing countries would be very useful for optimization of road traffic.

Therefore, the aim of this research is to reveal Indian taxi drivers' business behavior using long-term probe data. In this paper**,** we targeted three representative cites in India, Delhi, Mumbai, Hyderabad, and analyzed it such as the relationship of origin and destination and the tendency of trip length, trip time, and route choice behavior. The rest of the paper is organized as follows. In the next section, we explained the data outline and how to process the data. In section 4, we showed the results. In the last section, we concluded the paper with the research findings.

2 METHODOROGY

2.1 Target Area

We targeted 3 cities, Delhi, Mumbai, and Hyderabad, which have big population and known as the metropolitan area in India. Figure 2 shows the area and population in these cities. We extracted the target cities from the whole India with a shapefile based on administrative boundaries. The areas were calculated in QGIS. These 3 cities are different in area and population, and this might cause the difference of traffic conditions in each city.

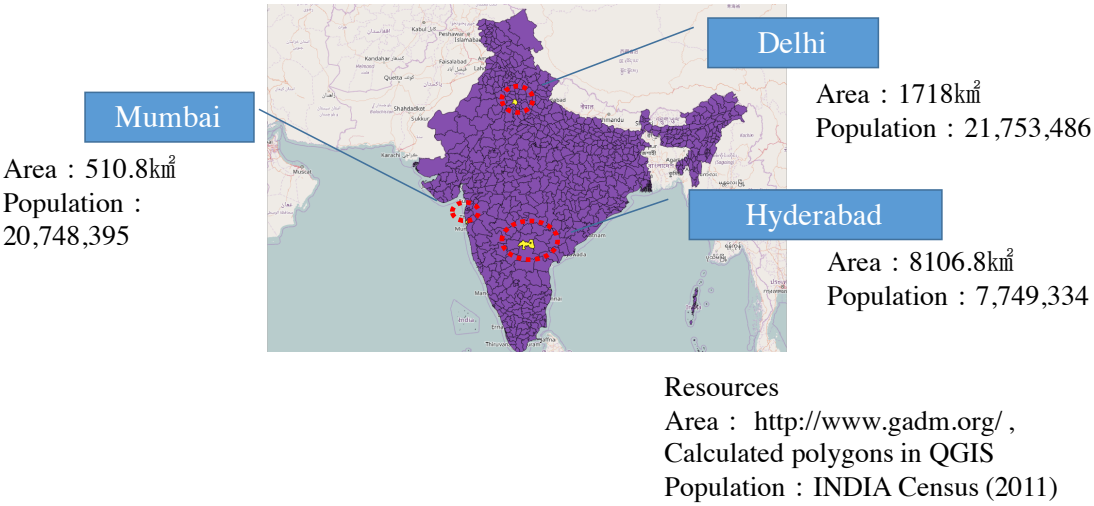

Figure 2. Basic information of 3 cities

2.2 Data Outline

Table 1 shows the outline of probe data. Probe data, or floating car data, means kinds of sensing car data, such as location, time and velocity. We used the taxi probe data for 14 months. Data is provided in every 1 minute through some process shown in Figure 3. These data cover whole India, but this time we picked up 3 cities, Delhi, Mumbai and Hyderabad as described. “Target” vehicle means the vehicle which has stayed in each area more than 1 hour at least once during the period. The border of the area follows one we got in this website: Global Administrative Area （2016） Table 2 shows the data format. they originally include the vehicle type, but it turned out to be incorrect, then we checked the vehicle type manually according to the vehicle register system (Ministry of Road Transport & Highways, 2017) and picked up taxi’s data.

Table 1. Probe data outline

Vehicle	Taxi (Delhi:651 cars, Mumbai:153 cars, Hyderabad:252 cars)
Duration	2015/4/1~2016/11/30（14 months）
Area	Delhi, Mumbai, Hyderabad
How to get data	Data is collected in the Comms server via GPRS/3G network and sent to Apps server via Internet
Data Interval	Approx.1 minute

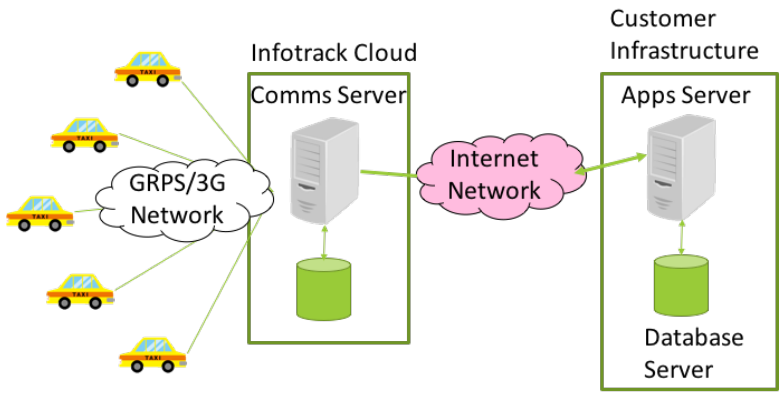

Figure 3. Process for data acquisition

Table 2. Data format

No.	Column	Example of Data
1	Vehicle No	VCH_KL-07-BV-3280
2	Track Time (YYYYMMDDHHMMSS)	2015-03-09 11:01:57
3	Latitude in Numeric No.	13.195316314697266
4	Longitude in Numeric No.	77.65283966064453

As a preparation for analysis, we made trip data based on each vehicle’s time and location. We extracted stay points from the probe data with outlier detection which Apichon W introduced (2013). We labeled vehicles which stayed within 200 m before next record as ‘STAY’ and others as ‘MOVE’**,** and also we gave each record road link id of OpenStreetMap according to latitude and longitude. This time we used the ids as of 2016/6/5. The data columns and examples are shown in Table 3**.**

Table 3. Trip data format

No.	Column	Example of Data
1	Vehicle No	VCH_KL-07-BV-3280
2	Date(YYYY-MM-DD)	2015-03-09
3	Sequence No	1
4	Mode label	STAY/MOVE
5	Distance	76241.9816
6	Duration	252.9833
7	Start time	12:51:00
8	End time	15:34:00
9	Point num	2
10	Trajectory (String with no1|datetime1|lat1|lon1|linkId1; no2|datetime2|lat2|lon2|linkId2 ...)	1|2015-04-01 16:36:06|12.990600|77.808000|353059; 2|2015-04-01 16:41:06|12.989300|77.785600|353057;
11	Date *duplicated with 2	2015-03-09

2.3 Process for POI matching

We extracted the origin and destination of each trip, and match POI (Point of Interest) for each. The data of POIs came from OpenStreetMap as of 2017/7/3. If all POIs are thought equally, it is possible to match inappropriate POI. For example, original POI data has both airport and shops in the airport. Hence, we divided all POIs into 4 categories such as airport, residential/commercial district, traffic base, others, and we did matching POI's category and name for each origin and destination following the priority as shown in Figure 4. "Commercial" district here means "Office" district based on the definition of OpenStreetMap.

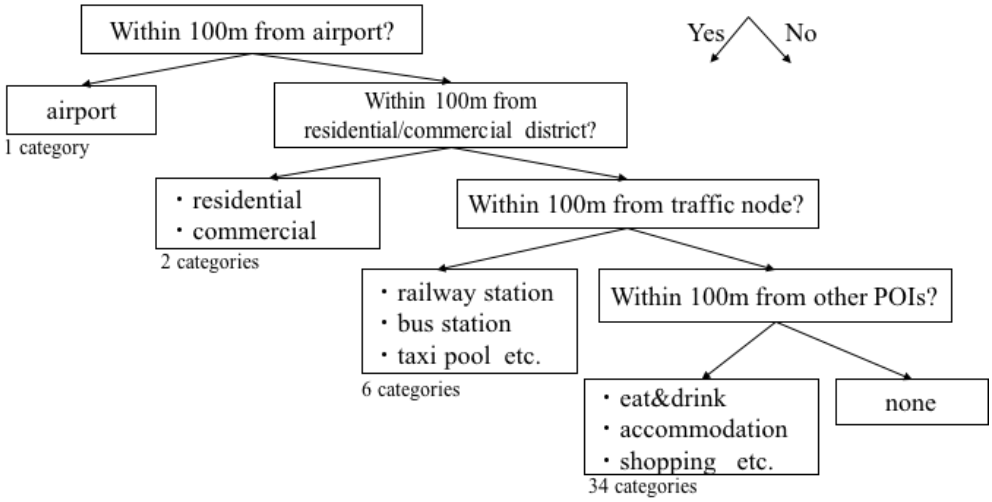

Figure 4. Process for matching POIs

After POI matching, we excluded the trips whose origin or destination were marked as "none" and analyzed the rest of the trip. Table 4 shows the target of the number of trips in each city.

Table 4. Number of trips

	Number of Trips		
	Total in 14 months	Daily Average	Average for Each Vehicle
Delhi	363,262	595.5	558
Mumbai	28,970	47.5	189.3
Hyderabad	81,484	133.6	323.3

3 TAXI'S BEHAVIOR

3.1 Characteristics of Origin and Destination

The result is that there are much more trips in all cities between residential and commercial district or between residential/commercial than trips between transportation base and restaurants. Table 5 shows the example of the ranking of the combination of origin and destination in Mumbai.

Table 5. Ranking of combination of origins and destination（Mumbai）

	Origin	Destination	No. of Trips
1	Commercial	Commercial	6,288
2	Commercial	Residential	2,485
3	Residential	Commercial	2,305
4	Residential	Residential	1,790
5	Eat & drink	Commercial	1,648
		Total number of trips：28970	

Looking at individual origin and destination, many trips started or finished to the nearest international airport in all 3 cities. Especially in Hyderabad, 35 % of all trips started the trip at or near Rajiv Gandhi International Airport, and many of their destinations were in residential areas located in 10~20km far from the airport (Figure 5). Based on the cheapness of taxi fare and low implement level of public transportation between airport and the town, the result suggests that taxi is popular transportation which connects major airport and city center.

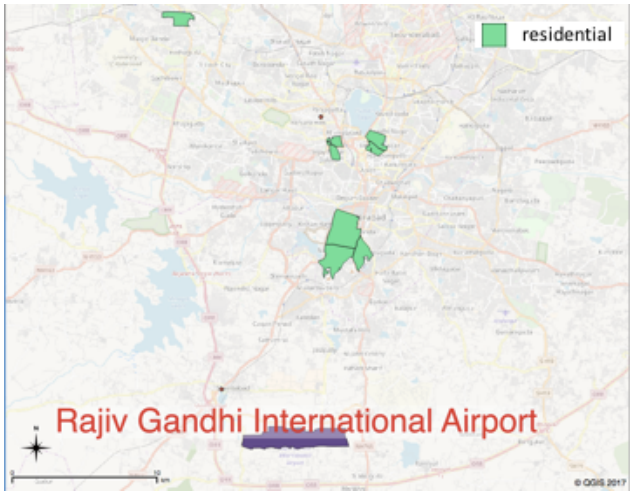

Figure 5. Rajiv Gandhi International Airport and major destinations

3.2 Route choice

We extracted the most number of the trip's origin and destination and analyzed trip time and length, route. These trips are in the suburb of Delhi, and their origin is in the residential area**,** and destination is in the commercial area. We analyzed 107 vehicles and 2700 trips. The shortest route for the trip is 3.5 km, and the red route as shown in Figure

6, and it takes about 8 minutes since the average traveling speed of taxi was 27 km/h. But in reality, although the mode of trip length is 3~4 km and trip time is 5~15 minutes, we can see some variety in terms of trip time and length as shown Figure 7,8.

Figure 6. Target OD

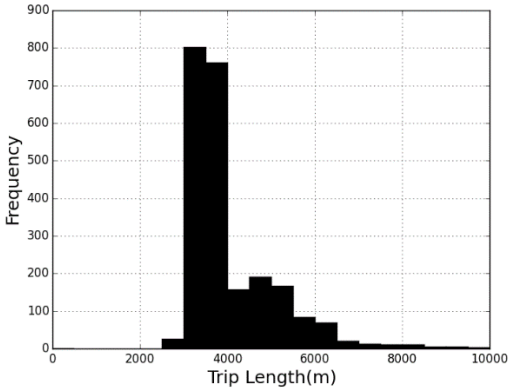

Figure 7. Histogram of trip length(N=2,700)

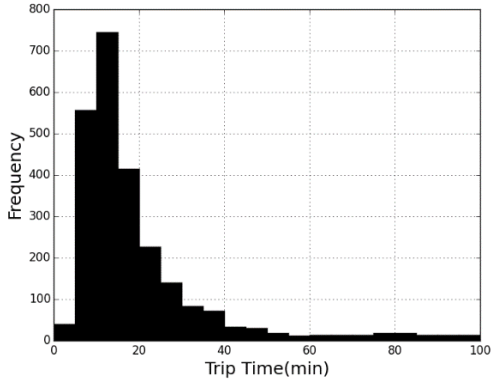

Figure 8. Histogram of trip time (N=2,700)

The trip has several routes, however**,** we found that there are few substitute routes which include minor road, and many trips are concentrated on the route which goes through major road which seems to be one of the causes of traffic congestion.

4 CONCLUSION

In this research, using long-term probe data in several Indian cities, we analyzed taxi behavior such as the attribute of origin and destination, and the tendency of trip length, time and route choice behavior. As a result, it is found that taxi is one of popular transportation for people living in these cities, and we also found substitute route would be very few and the trip is concentrated on major road.

However, we have two weak points in this analysis. First, we couldn't get the data whether the taxi has customer or not which might have influenced the drivers' route choice behavior. For instance, when the driver has a customer, they are likely to choose the shortest or fastest route, but on the other hand when the taxi doesn't have a customer, they tend to drive more randomly because they want to get a customer. So we need to be careful to analyze the result. Second, the priority of POI matching seems to be insufficiently thought. We need to think the priority and threshold more carefully. Being careful for these points, we will continue to do analysis of environmental impact on traffic volume and route choice behavior.

ACKNOWLEDGE
Obtaining probe data described in this paper was supported by Ltd. Zenrin Data Com.

REFERENCE

Aparajita C, Sudakshina G, 2014. Traffic congestion in the metropolitan city of kolkata. Journal of Infrastructure Development vol. 6-1 43-59.

Apichon W, Teerayut H, Natsumi O, Yoshihide S, Ryosuke S, 2013. Trip reconstruction and transportation mode extraction on low data rate GPS data from mobile phone. CUPUM2013.

Bhuvanachithra C, Dimitrios Z, 2012. Congestion mitigation measure in hyderabad– a midnight summer dream?. South Asia Chronicle S. 58-92.

Robert Hijmans. Global Administrative Areas. Retrieved August 9, 2017, from http://www.gadm.org/

Government of National Capital Territory of Delhi, 2011. Delhi statistical handbook.

Hiroaki K, Hiromitsu Hattori, 2015. Analysing and modeling individualities of taxi drivers' driving behaviors using probe-car data. The 29th Annual Conference of the Japanese Society for Artificial Intelligence.

India stats: million plus cities in India as per census 2011. Retrieved March 17, 2017 from http://pibmumbai.gov.in/scripts/detail.asp?releaseId=E2011IS3

Ministry of Internal Affairs and Communications Bureau of Statistics in Japan, 2017World's Census.

M Absar A., Faisal A, 2013. Urban transport systems and congestion: a case study of Indian cities. Transport and Communications Bulletin for Asia and the Pacific no.82.

P.S.Bhargavi, N. Kannaiya Raja, 2011. A study on the effective traffic management of roads to reduce road accidents in Tamilnadu. International Journal of Enterprise Innovation Management Studies, vol2, no2. 172-190

Rameshwar, D. , Sandeep, J. ,Kewal, S, 2011. Growth rate of motor vehicles in India -impact of demographic and economic development. Journal of Economic and Social Studies, vol1, no2.

Toshio Y, Daisuke F, Ryuichi K, 2005. An analysis of vacant taxi drivers' behavior based on probe data. The 31th Annual Conference of the Japanese Infrastructure Planning and Management

Trip Adviser, 2011. Taxi Fare in Major Cities in the world, Retrieved August 9, 2017, from http://tg.tripadvisor.jp/TaxiFare/

Ministry of Road Transport & Highways. National Register e-Services of Registered Vehicles, Retrieved June 7, 2017, from https://vahan.nic.in/nrservices/