

Transportation Melting Pot Dhaka: Road-link Based Traffic Volume Estimation from Sparse CDR Data

Yoko Hasegawa¹, Yoshihide Sekimoto¹, Takehiro Kashiyama¹, Hiroshi Kanasugi²

¹Institute of Industrial Science, University of Tokyo

² Earth Observation Data Integration and Fusion Research Initiative
Ce507, 4-6-1, Komaba, Meguro-ku, Tokyo 153-8505, Japan
Tel: +81-3-5452-6406

(yokohase, sekimoto, ksym, yok)@iis.u-tokyo.ac.jp

ABSTRACT

Understanding traffic conditions in urban areas is an important research direction, especially in rapidly growing cities that still struggle with congestion and inefficient traffic control strategies. The purpose of this study is to estimate the road-link scale traffic conditions of the metropolitan area of Dhaka, where a variety of transportation demands get mixed up as if it were a ‘melting pot’, based on sparse mobile phone data and road networks. The mobile phone data used here is the Call Detail Records (CDR). Our method extracted Origin and Destination (OD) from CDR in two ways. One is a simple extraction of continuous records with base station differences, and another is an extraction of trips between significant locations through CDR clustering and trip segmentation. Full-day link traffic volume is then estimated by assigning hourly trips with each OD to routes on actual road network. The methodology is demonstrated using 1 month CDR from 6.85 million users of Dhaka, with only 5.8 logs per day in average. Our estimation results show a relatively strong correlation ($r=0.75$) with the actual traffic count in a road-link scale. Moreover, the fact that the estimation results have close accuracy with the Person Trip survey data based estimation suggests that traffic conditions understanding based on long-term mobile phone data is a valid method for large-scale traffic survey.

Categories and Subject Descriptors

H.2.8 [Database Applications]: Spatial databases and GIS
J.4 [Social and Behavioral Sciences]: Sociology

General Terms

Algorithms, Experimentation, Verification

Keywords

Traffic conditions estimation, Call Detail Records (CDR), Dhaka

1. INTRODUCTION

Understanding the dynamic change in distribution and flow of traffic has become a necessary factor for strategic traffic network management in urban areas. Detailed and up-to-date information of the current traffic conditions is needed for efficient action,

Permission to make digital or hard copies of all or part 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. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Urb-IoT 2014, October 27-28, Rome, Italy
Copyright © 2014 ICST 978-1-63190-037-2
DOI 10.4108/icst.urb-iot.2014.257272

especially in expanding cities with traffic networks still under rapid development.

Many researchers are now focusing on understanding human mobility features using mobile phone data [3,4,6,7]. Analysis of user-anonymized mobile phone billing records, including the Call Detail Records (CDR), have an especially high potential for effective traffic conditions estimation, due to their wide population and area coverage. Another benefit of using this data is that there is no need for additional infrastructure, because mobile phones have already become one of the most important lifelines in many countries.

However, Wang *et al.*[7] only recently began to study more detailed estimations of traffic flow and distribution on a road-link scale, where they estimated San Francisco and Boston traffic with relatively high accuracy. Iqbal *et al.*[4] have tried similar method with Dhaka CDR, but its applicability is only shown in simple road network among 67 nodes. The validation of detailed estimation of traffic conditions on actual road network based on sparse CDR data is left to be demonstrated in developing countries.

Another feature of this work is that we take in consideration an important feature of traffic conditions in metropolitan area of a developing country, which is that wide variety of transportation demands are mixed up to make disordered and heavily congested situation. For example, in Dhaka, more than 20 vehicle types, with wide range of number of passengers of 1 to 100, could be observed [5]. The novel approach we take for this is converting population-based traffic volume to vehicle-based traffic volume.

The final purpose of our work is to estimate the detailed traffic conditions of the Dhaka metropolitan area (DMA) at the road-link scale, using sparse 1 month CDR of 6.85 million users, which has 5.8 logs per day per user in average, as basic trip information. The remainder of this paper is structured as follows. Section 2 explains two methods for extraction of possible ODs from CDR and how hourly trip production volume is assigned to each ODs. Next, section 3 presents results of a comparison of our traffic conditions estimation method based on the CDR of Dhaka, with the traffic survey results. Section 4 discusses the main causes of errors and future research. Finally, section 5 presents our conclusions.

2. METHODOLOGY

We explain the overall outline of our method for road-link scale traffic conditions estimation in this section. First, we extracted possible OD from CDR based on two different trip definitions as follows. The method flows for both definitions are shown in Figure 1.

(i) Transient origin destination (t-OD)

We extracted transient origin destination (t-OD), as defined by Wang *et al* [7]. A trip is defined when a same mobile phone user is observed in two distinct Voronoi zones within one hour. Wide

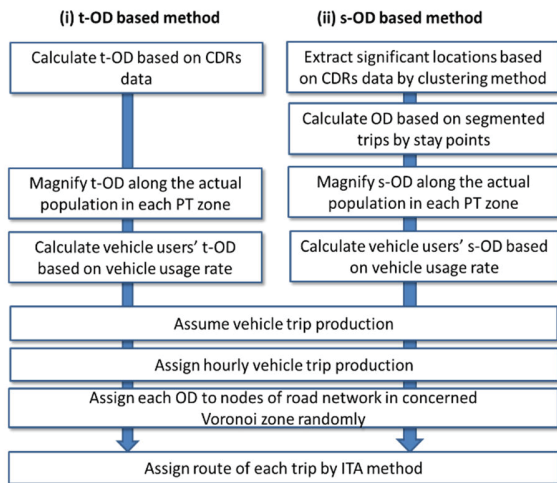


Figure 1. Link traffic volume estimation outline

distribution of travel demands could be captured from the long-term, mass distinct users' dataset. However, some parts of the entire dataset will be disregarded when there are more than an hour of blank between two continuous records.

(ii) Origin-destination limited to significant locations (s-OD)

Significant locations for each user were defined as representative points of clusters derived from long-term CDR. The mean-shift procedure [1] was applied for spatial clustering.

Then, a trip was defined when the same mobile phone user was observed in two distinct clusters within one hour. The Voronoi zones in which the representative points of clusters were included were extracted as the s-OD for the trip.

We magnified the counts of each OD based on the residential population statistics derived from the census data, comparing with the identified home place of the users. This is to adjust the bias in the mobile phone users' distribution. Also, to assure that we assigned only the trips attributed to vehicles, we multiplied the vehicle usage rate, which is constant for each Voronoi zone.

Then, we assigned hourly trip numbers to these ODs based on the calculated distribution of the ODs. The total number of hourly trips was calculated from the census data. To ensure that we only considered vehicle trips on the road network, the person based trip count was adjusted to the vehicle based trip count, by dividing the hourly trip number by the average passenger per vehicle (PPV) for that hour. PPV was calculated as the sum of the product of the modal share rate of the vehicle type and the average PPV of that type.

Finally, each vehicle trip assigned to ODs at this point was assigned to routes between the OD and road-link traffic volumes were calculated.

3. RESULTS

We estimated the routes for each OD pair of vehicle trips as the final step in the process. We used the incremental traffic assignment method (ITA method) [7] to assign the routes. ITA is a relevant method to assign a large number of trips to road networks with relatively small computational complexity. The original OD was first split into four small sets, which contained 40%, 30%, 20% and 10% of the total number of trips distributed. Then, the trips in the first set were assigned using the free travel time along the routes computed by Dijkstra's algorithm [2]. The actual travel time was assumed from the Bureau of Public Roads (BPR) function (1) to update the travel time. The trips in the second set were assigned using the updated travel time.

$$t_a = t_f(1 + \alpha(VOC)^\beta) \quad \alpha = 0.15, \beta = 4$$

t_f : free travel time

t_a : updated actual travel time

VOC : vehicle operation cost

(link traffic volume / road capacity)

(1)

Iteratively, all the trips in the four sets were assigned to routes between OD. Before this route assignment procedure, a node on the road network inside each Voronoi zone was randomly selected for the beginning and end of the route, because our estimated ODs are in Voronoi zone scale.

We obtained the hourly traffic volume for each road-link as a result of the route assignment of the trips to the road network. The total road-link traffic volume was validated by comparing it with the 24 hour traffic count of 40 distributed locations from the 2009 JICA observation survey. Figure 2 shows the total estimated road-link traffic volume based on the s-OD method and the JICA survey points. Figure 3 shows a comparison of the correlation with the observed traffic count with results of t-OD based estimation (Figure 3a) and s-OD based estimation (Figure 3b). For comparison, Figure 3c shows the correlation with the estimation result based on the OD matrices from Person Trip (PT) survey held also by JICA in 2009.

4. DISCUSSION

The estimation based on the ODs extracted from CDR were quite close in accuracy to the PT OD based estimation, however the correlation coefficient was around 0.75 (r) and still requires further development. This suggests a strong potential for traffic condition understanding based on long-term mobile phone data to be one valid alternative to large-scale traffic surveys. Comparison of the two methods for ODs extraction from CDR shows that the total counts of all the road-links was quite small, including about 1/3 of the observation counts. The t-OD distance distribution (Figure 4) that most of the extracted trips based on t-OD were very short in distance compared to the PT ODs, such as trips between neighboring Voronoi zones. This is likely to be extracted in areas with a high density of base stations, where the actual 3 ~ 4 km distance trip could be divided into 4 or 5 short distance t-OD trips. The results from the s-OD based method show that if we consider

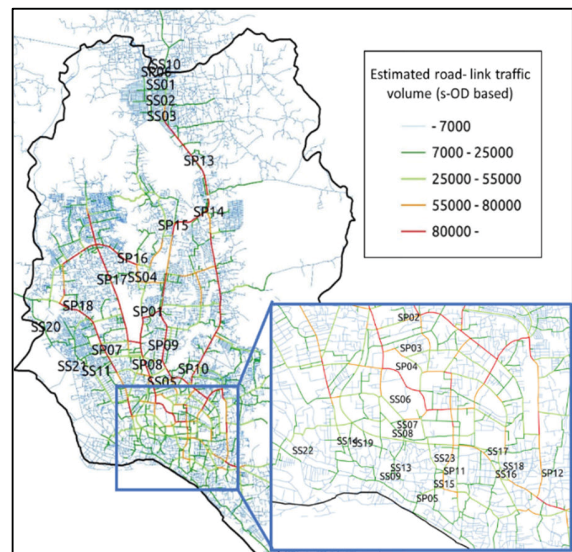


Figure 2. Estimated 24h road-link traffic volume calculated based on s-OD and JICA survey points..

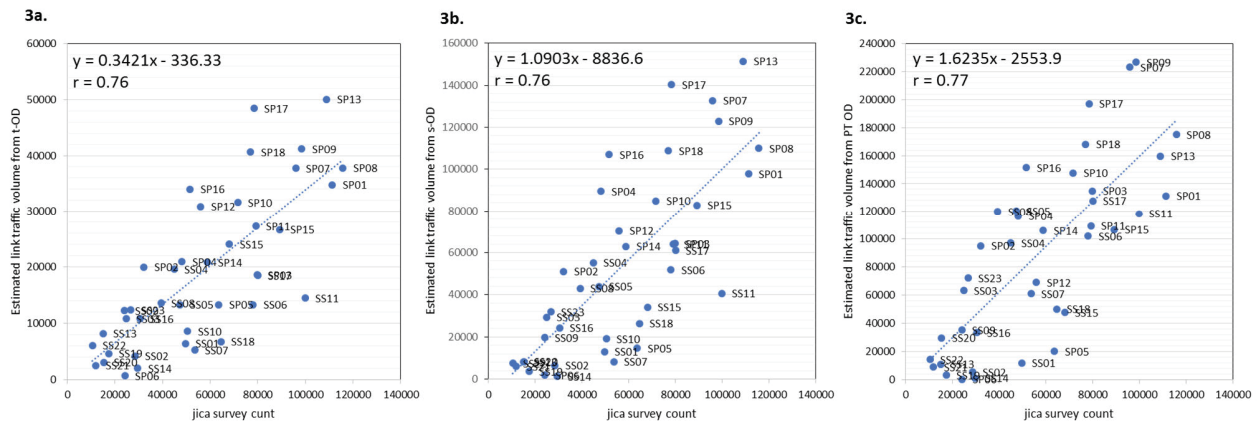


Figure 3. Correlation with observed traffic count with estimated link traffic volume
(3a) t-OD based estimation (3b) s-OD based estimation (3c) PT OD (distributed to voronoi zone scale) based estimation

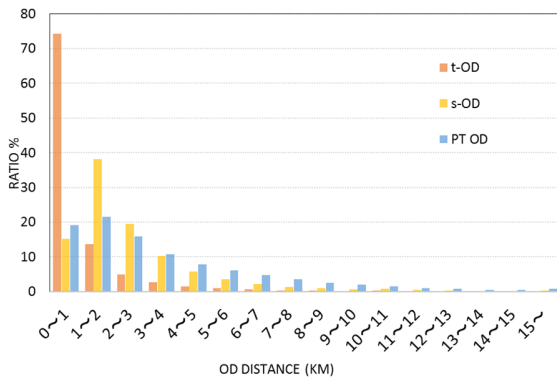


Figure 4. OD distance distribution

the small spatial differences between the continuous CDR records as stay actions, the extracted trips led to better simulation of the actual traffic demand than did those from the t-OD based method. The factors listed below should be considered as possible sources of error regarding the accuracy of our estimation.

(1) Trips between inside and outside the DMA

The ODs extracted from CDR in this study were limited to trips inside the DMA, although, according to the JICA report, about 2% of the total actual trips are between inside and outside the DMA. Figure 9 shows that the survey points close to the borderline of the DMA (SP06, SS10, SS01, SS02, and SP05) are affected by this limitation, showing smaller estimated traffic volume than the observation survey count.

(2) 3 year time lag between the CDR and PT survey

The CDR data used in this work was recorded 3 years after the 2009 PT survey, therefore, considering the unparalleled pace of growth in the DMA, there could have been major changes in the structure of the city within that time. This includes expansion of the city's area, development of the road network, etc., and these changes may not be reflected in the data used for validation.

(3) Parameter settings for route assignment

Figure 9 shows that survey points labeled SP (Primary road) show excessive amounts of estimated road-link traffic volume, while points labeled SS (Secondary road) show the opposite tendency. The examined parameter s for route assignment may have resulted in assigning too many trips to the limited primary road-links therefore various parameter settings should be tested for more reliable estimation.

5. CONCLUSION

In this paper, we estimated the road-link scale traffic conditions of Dhaka using 1 month CDR from 6.85 million users. The methods used were mainly developed by Wang *et al.*, and the purpose of this work was to test the applicability of this method to actual road network of a developing city, based on sparse mobile phone data. Our estimation had a relatively high correlation coefficient ($r = 0.75$) with the actual road-link scale traffic count. Moreover, the estimated road-link traffic count distributions based on the mobile phone data showed close accuracy with the estimations based on PT survey data. We conclude from our findings that with long-term mobile phone data and road network data, valid modeling of traffic demand can be done in cities with various traffic features.

We plan further development to improve the accuracy of the method, by addressing the list of possible error factors with detailed validation in future research.

6. ACKNOWLEDGMENTS

We thank GrameenPhone for providing Dhaka CDR dataset.

7. REFERENCES

- [1] Cheng, Y. 1995. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 17, 8 (1995).
- [2] Dijkstra, E.W. 1959. A note on two problems in connexion with graphs. *Numerische Mathematik*. 1, 1 (1959), 269–271.
- [3] González, M.C. et al. 2008. Understanding individual human mobility patterns. *Nature*. 453, 7196 (Jun. 2008), 779–82.
- [4] Iqbal, M.S. et al. 2014. Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies*. 40, (Mar. 2014), 63–74.
- [5] JICA: Preliminary Survey for Dhaka Mass Rapid Transit Development Project Final Report (2010)
- [6] Ratti, C. et al. 2006. Mobile Landscapes: using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design*.
- [7] Wang, P. et al. 2012. Understanding road usage patterns in urban areas. *Scientific reports*. 2, (Jan. 2012), 1001.