DEVELOPMENT OF PEOPLE MOVEMENT MODEL USING DYNAMIC ESTIMATED POPULATION DATA

Mayumi Hadano* [1], Yuki Akiyama [2], Ryosuke Shibasaki [3], Toshikazu Nakamura [4]

[1] Master's degree student, Graduate School of Frontier Science, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan; Tel: +81-4-7136-4307; E-mail: hadano@csis.u-tokyo.ac.jp
[2] Researcher, Center for Spatial Information Science, The University of Tokyo, 4-6-1, Komaba Meguro-ku, Tokyo 153-8505, Japan; Tel: +81-3-5452-6417; E-mail: aki@iis.u-tokyo.ac.jp
[3] Professor, Center for Spatial Information Science, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan; Tel: +81-4-7136-4290; E-mail: shiba@csis.u-tokyo.ac.jp
[4] Ph.D. student, Graduate School of Frontier Science, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan; Tel: +81-4-7136-4307; E-mail: ki_ki_gu@csis.u-tokyo.ac.jp

KEY WORDS: People Movement, Dynamic Population, Daytime Population, Traffic Planning

ABSTRACT: People movements are often investigated with person trip survey, OD survey, and a questionnaire survey. Limitations of these methods are taking much time and money and low sampling rates. In recent years, more detailed human movements can be monitored by developments of mobile devices and GPS technologies. However there are only few examples of applying mobile technologies on a large scale because of these costs to monitor many people. On the other hand, "dynamic population data" become available. It is population data with high spatio-temporal resolution estimated by aggregating the log data of mobile phones and other methods. However the data represents population distribution and not how people move. The purpose of this study is to develop a method to estimate people movement from hourly and daily dynamic population data. We made a detailed model which can describe human movement with every hour interval. Estimated people movement data will be expected to apply to fields of transportation planning, disaster prevention and business.

1. INTRODUCTION

Monitoring people movements has become important in fields of transportation planning, disaster prevention, and business. People movements are often investigated with Origin-Destination (OD) survey, questionnaire survey such as person trip survey and so on. Among these surveys, the person trip survey is the most largely studied in order to analyze urban transportation and monitor microscopic people flow. However, the person trip survey takes much time and money and its sampling rates are low (about 3 percent). Also frequency of person trip survey in each metropolitan area is about 10 years. In recent years, more detailed human movements can be monitored by developments of mobile devices and GPS technologies. For example, Ratti *et al.* (2006) tracked the number of stationary people from the number of registered mobile phones at a base station and gained insight into complex and rapidly changing urban dynamics phenomena. Caceres *et al.* (2007) also analyzed mobile phone location data in order to infer traffic data such as journey Origin-Destination matrices or traffic counts at given points in the road network. However there are only few examples of applying mobile technologies on a large scale because of these costs to monitor many people. On the other hand, some "dynamic population data" such as "Agoop data" become available. It is population data with high spatio-temporal resolution estimated by aggregating the log data of mobile phones and other methods. However the data represents population distribution and does not monitor people move. The purpose of this study is to develop the method to estimate people movement using means Origin-Destination estimation from hourly and daily dynamic population data.

2. DATA DEVELOPMENT

In this study, we applied Agoop data (described in following text) as the dynamic estimated population data. For OD estimation, we used the movement probability of Person trip data (PT data) processed from person trip survey data with calculating initial value of OD matrix.

.

2.1 Study area and aggregate unit

A study area is Sendai metropolitan area in Miyagi prefecture, Japan (Figure 1). Miyagi prefecture conducted person trip survey at this area in 2002. Sendai city is the largest city in Tohoku region.
Aggregate units are the 500m-square grids the called Japanese standard regional mesh. Many statistical grid data of Japan are aggregated into this grid and load data in this study is also aggregated into this grid. National census data

and the Agoop data are also aggregate into this 500m-square grids.

2.2 Person trip survey and PT data

We used the Person trip data (PT data) processed from person trip survey data. Miyaki prefecture conducted this survey in October 2002. This season is regarded as best season for explorations of person flow (Hukumoto and Nakamura, 2011). The coverage area of this survey was Sendai metropolitan area including 13 cities. The number of people in the sample was about 90,000.

In the person trip survey, respondents were required to fill entries on each visited place and each trip between two places. In addition, they were also required to answer basic individual information such as gender, age, and occupation. Because of privacy considerations, the places are recorded as a rough address to avoid specifying exact locations. Each trip must include a departure time, arrival time, and purpose.

PT data was reconstructed from this person trip survey data by the “People Flow Project” in Center for Spatial Information Science, The University of Tokyo. We use this PT data at 8:00 and 9:00 to estimate people movement at commuting time. The method of reconstruction is described by Sekimoto *et al.*(in press).

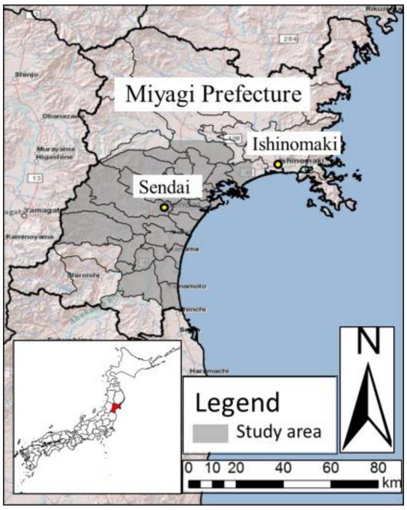

Figure1: Study area

2.3 Agoop data

We used the Agoop data in 2009 (AGOOP Corp.) as distribution data of dynamic population aggregated into 500m-square grids. This data enables to monitor the dynamic populations in each month and each hour. The estimated dynamic population means the average hourly number of people located in each grid. Figure 2 shows a grid map of population data in study area. The national census data has only static population such as resident population or working population, while Agoop data can describe dynamically changing population that includes shopping population and recreation population. This data was made by investigating facilities gathering lots of people and defining the pattern of population distribution change in time series at each kind of facilities. In this study we used the Agoop data between 8:00 and 9:00 on weekdays in October. Figure 2 shows the grid map of dynamic population data at 8:00 and 9:00.

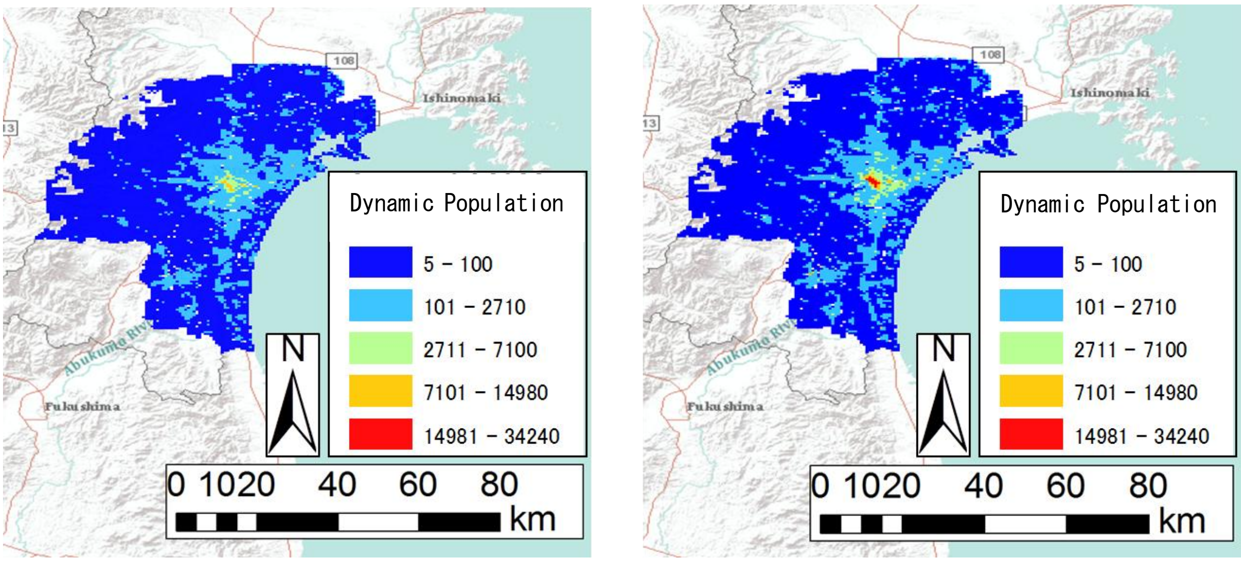

Figure2: Grid map of dynamic population data at 8:00 (left) and 9:00 (right).

2.4 Correlation between PT data/Agoop data and national census data

Because PT data and Agoop data are estimated data, we need compare them to the existing statistical data about population. Therefore we analyzed the correlation between them and national census data in 2005. Figure 3 shows the correlation between the resident population of the national census in 2005 and PT data/Agoop data at 3:00 in each grid. The correlation coefficient between PT data with the national census was 0.54 and the correlation coefficient between Agoop data and the national census was 0.92. Therefore both data has correlation to national census data.

2.5 OD estimation

Whole data processing flow of OD estimation is given in Figure 4.

(a) Setting the initial value of OD matrix

First using PT data we aggregated the location i.e. longitude and latitude of people into 500m-square grids. Then we made OD matrix between 8:00 and 9:00 using personal ID. The initial movement probability $Prob_0$ (i, j) is calculated as demonstrated in Equation 1:

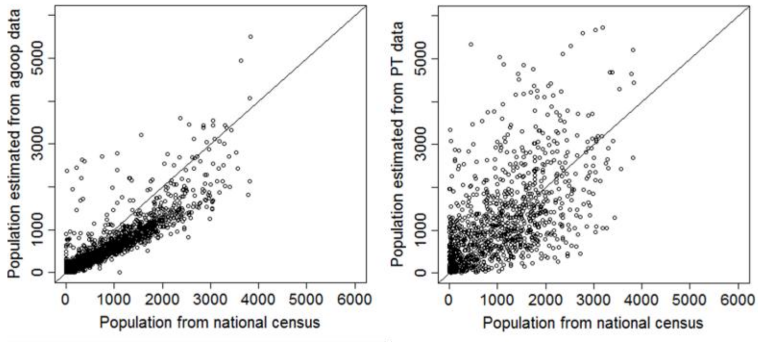

Figure 3. The correlation between the resident population from the national census of 2005 and Agoop (left)/PT data (right) at 3:00 in each grid

$$Prob_0(i,j) = \frac{P(i,j)}{\sum_j P(i,j)} \quad (1)$$

where $Prob_0$ (i, j) is the probability that person of PT data located at mesh i at 8:00 moved to mesh j at 9:00, $P(i,j)$ composes OD matrix and means the number of people that located at mesh i at 8:00 moved to mesh j at 9:00. Finally the initial values of OD matrix are defined as Equation 2:

$$P_0(i,j) = Prob(i,j) \times Pop(t,i) \quad (2)$$

where $Pop(t,i)$ means the number of people of Agoop data located at mesh i at time t (=8:00).

(b) Adding up the estimated OD matrix and calculating residuals

We added up the estimated results of OD matrix at each destination grid. The residuals between them and the number of people that located at mesh j at i+1(=9:00) were calculated as demonstrated in Equation 3:

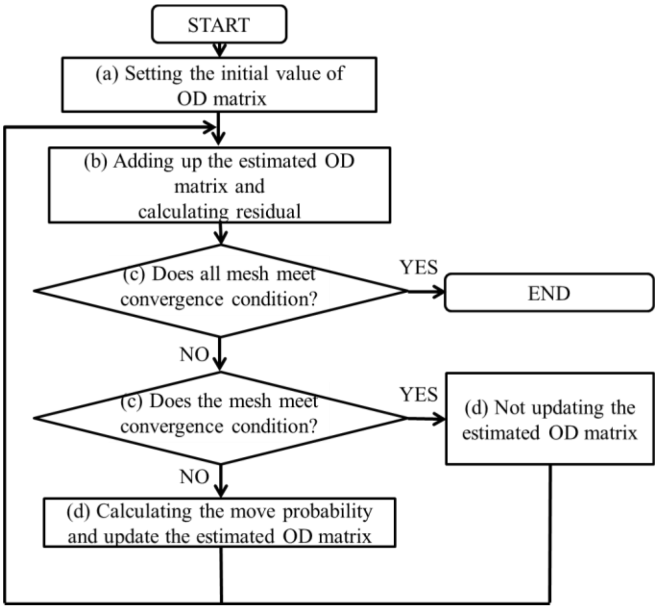

Figure 4. Whole data processing flow of OD estimation

$$\Delta P_n(j) = \sum_i P_n(i,j) - Pop(t+1,j) \quad (3)$$

(c) Convergence condition test

Using the values of residuals, convergence condition test was done at each grid. We used the following condition expressed by Equation 4.

$$\Delta P^2{}_n(j) < 1 \quad (4)$$

OD estimation finished when all grids satisfy this condition.

(d) Calculating the movement probability and update the estimated OD matirix

The new movement probability between 8:00 and 9:00 is calculated as demonstrated in Equation 5.

$$Prob_{n+1}(i,j) = \frac{P_n(i,j)}{\sum_j P_n(i,j)} \quad (5)$$

The grids satisfying the convergence condition were not updated (Equation 6) and the grids not satisfying the condition were update using Equation 7:

$$P_{n+1}(i,j) = P_n(i,j) \quad (6)$$

$$P_{n+1}(i,j) = \begin{cases} P_n(i,j) + Adj_n(i,j) & (\text{when } P_n(i,j) > Adj_n(i,j)) \\ 0 & (\text{when } P_n(i,j) \le Adj_n(i,j)) \end{cases} \quad (7)$$

where Adj_n (i,j) was defined as Equation 8.

$$Adj_n(i,j) = Prob_{n+1}(i,j) \times \Delta P_n(j) \quad (8)$$

OD matrix was estimated by repeating the process between (b) and (d).

3. RESULTS OF OD ESTIMATION

Figure 5 shows the result of OD estimation. The vector in this figure means people movements. The initial point of vector is the center point of the grid where people are located at 8:00 and the direction of vector means average direction that people moved between 8:00 and 9:00. The length of vector is fixed. In addition, the color of the vector means the number of moving people excluding stopping people and red color expresses many people movements. Figure 6 shows the OD matrix of PT data with the same method. Figure 7 and 8 are zoom-in figure to Sendai city from figure 5 and 6, respectively. Although both figures grasp the people flow that move from suburbs to Sendai city at commuter time, details of people flows was different. For example in the result of OD estimation there are a lot of moving people at a distance of the central area of Sendai city, while in PT data there are many moving people at the central area of Sendai city.
The coefficient between the result of OD estimation and PT data was 0.14 and there is low correlation.

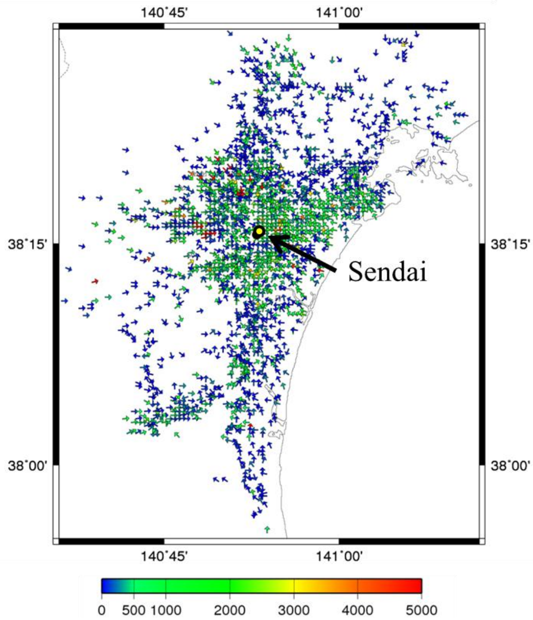

Figure 5. The result of OD estimation using Agoop data between 8:00 and 9:00..

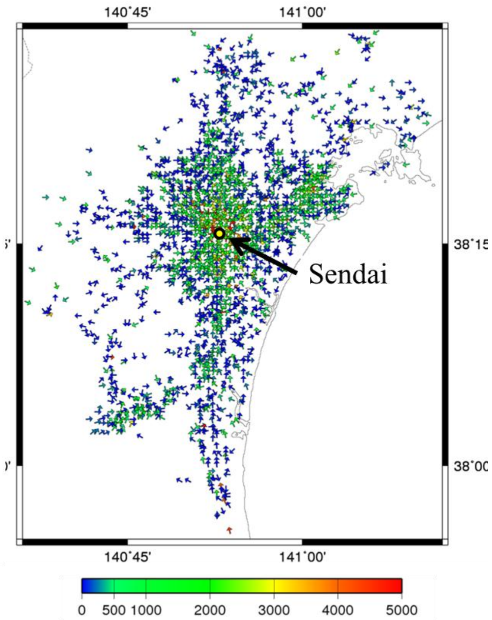

Figure 6. Visualized OD matrix of PT data between 8:00 and 9:00.

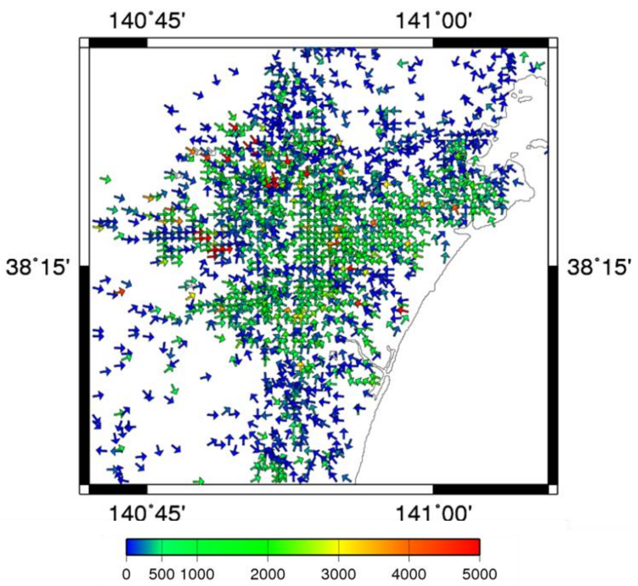

Figure7. Zoom-in figure to Sendai city from figure 5.

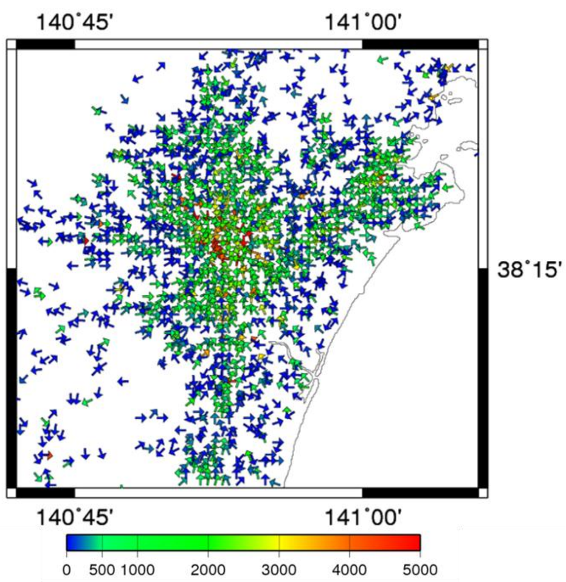

Figure 8. Zoom-in figure to Sendai city from figure 6.

4. CONCLUSION

In this paper, we developed the method to estimate people movement from hourly dynamic population data using the Agoop data and visualized the people flow with vector in Sendai metropolitan area. We applied the movement probability of PT data to set the initial OD matrix and conducted convergence condition test.

For future work, OD estimation will be improved if the way of setting the initial OD matrix. In this study, we used the PT data for setting the initial OD matrix however this data has spatial bias especially in suburb area. In the PT data, the locations of people are recorded by the survey zone that is defined by the local government. Because the survey zones of suburb areas are generally larger than that of urban areas. In addition, because PT data is a sampling data, PT data has smaller investigation scope than the national census. Figure 9 shows the area where people exit at 3:00 in PT data and the resident population data in national census. The range of the national census data is larger than that of PT data. In this study we used the PT data for setting the initial OD data and ignore the grids that is not included in PT data, the result of OD estimation has also spatial bias shown in Figure 5 and 7. In addition, the areas where person trip survey is conducted are limited. Improved estimated people movement data will be expected to apply to fields of transportation planning, disaster prevention and business.

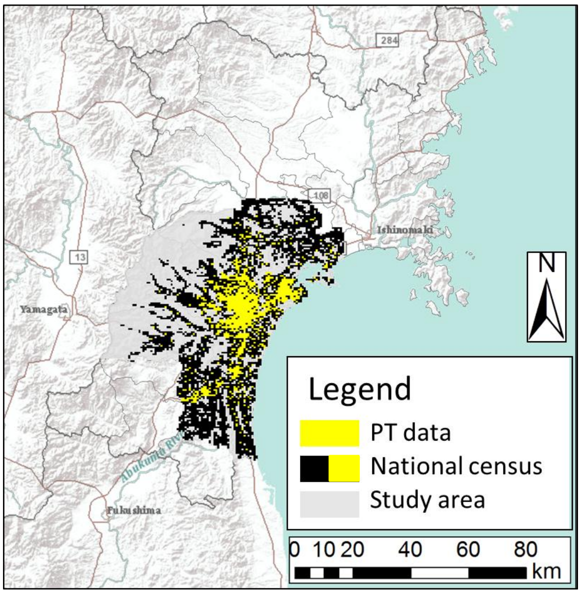

Figure 9. The areas where people exit at 3:00 in PT data (yellow) and the range of resident population data in national census in 2005 (yellow and black).

ACKNOWLEDGMENT
We are contributed Agoop data by Agoop Corp. We would like to thank Agoop Corp..

REFERNCES

References from Journals:

Caceres N, Wideberg, J.P, and Benitez, F.G, 2007. Deriving origin-destination data from a mobile phone network. IET Intelligent Transport Systems, 1 (1), pp. 15-26.

Hukumoto, H. and Nakamura, Y., 2011, The 5th nationwide person trip survey in 2010. Traffic Engineering, 46(2), pp. 12-15.

Ratti, C., Pulselli, R.M., Williams, S., and Frenchman, D, 2006. Mobile landscapes: using location data from cell-phones for urban analysis, environment and Planning B: Planning and Design, 33 (5), pp. 727-748.

Sekimoto, Y., Shibasaki, R., Kanasugi, H., Usui, T. and Shimazaki, Y., 2011. PFLOW: Reconstruction of people flow by recycling large-scale fragmentary social survey data. IEEE Pervasive Computing, 10 (4) (printing).