Human mobility decline in disaster analyzed by large-scale positioning data

Makoto Okumura[1], Jinyoung Kim[2], Kento Bunphong[3]

1) Tohoku University, RIEC No.2 Buld. 2-1-1 Katahira, Aoba-ku, Sendai, 980-8577,JAPAN, phone/fax: 81-22-217-6369, mokmr@m.tohoku.ac.jp
2) Tohoku University, RIEC No.2 Buld. 2-1-1 Katahira, Aoba-ku, Sendai, 980-8577,JAPAN, phone/fax: 81-22-217-6382, kim@irides.tohoku.ac.jp
3) Nippon Telegraph and Telephone East Corporation, bunpon@cneas.tohoku.ac.jp

Abstract

The Great East Japan Earthquake caused a long-term human mobility decline in north-eastern Japan including Sendai city, and people who had lived in the suburb of Sendai city had difficulty in coming to the downtown. In this study, we assume that the mobility decline possibly affects the change of hourly location distribution on population. As the relationship of location distribution between before and after the earthquake is clarified, consequently the restoration process from mobility change can be understood quantitatively.

Keywords

Big data, Mobility, Urban Structure, Disaster, Recovery

1. Introduction

The Great East Japan Earthquake caused a short- and long-term human mobility decline in north-eastern Japan including Sendai city. Gasoline shortage spread all over the affected area for about a month since the earthquake occured[1]. This gasoline shortage was one of the reasons of the daily mobility decline, although the public transportation service (e. g. subway, bus, etc.) was restarted in the early stage comparatively. However, the detailed causes of mobility decline haven't been analyzed yet.

In this study, we assume that the mobility decline possibly affects the change of hourly location distribution of population. And the population location distribution data are obtained from GPS positioning data. As the relationship of the location distribution between before and after the earthquake is clarified, we can evaluate the mobility decline quantitatively. Moreover population distribution can be estimated in a certain period. Consequently, the restoration process from the mobility change can be understood quantitatively.

2. Data processing

2.1. Large-scale positioning data

The research area is Sendai city where was affected by the Great East Japan Earthquake on 11 March 2011. The period of this research is set 61 days, from 1th March 2011 to 30th April 2011, and the period is divided into two, before and after the earthquake.

NTT DOCOMO which is the largest mobile service company in Japan provides the map service via about 800,000 GPS mobile phones, and positioning data are collected from the mobile phones automatically. The data are aggregated to a mesh hourly through statistical processing not to enable the specific individual to be identified, and then, they are processed and expanded as the mesh-unit population.

2.2. Mesh-unit population rate

The length of one side of a mesh is about 250 meters (one fourth of regional mesh). There are 12,904 meshes in Sendai city.

We suppose the population distribution is the consequence of mobility in this study. However, the number of population processed from GPS positing data can't be used as it is because of missing data after the earthquake. Instead, the rate of population is used for interpreting the mobility. The rate of population and the distribution of the population rate are defined using the mesh-unit population data as the following equation (1);

$$p_{d,t}^{m} = {n_{d,t}^{m}} \Big/ {\sum_{m=1}^{M} n_{d,t}^{m}} \qquad (1)$$

$$P_{d,t} = \{p_{d,t}^{1}, \cdots, p_{d,t}^{M}\}$$

where,

M : the total number of mesh in research area

$p_{d,t}^{m}$: the rate of population in mesh *m* on the date *d* the during time interval *t*

$n_{d,t}^{m}$: the number of population in mesh *m* on the date *d* the during time interval *t*

$P_{d,t}$: distribution of population rate in research area on the date *d* the during time interval *t*

2.3. Determination of the data missing

A cloud of people used mobile and fixed-line phones to confirm their family's and/or friends' safety after the Great East Japan Earthquake, and mobile phone couldn't be connected for a while due to the surge of telecommunication traffic. Some of mobile base stations were stricken by the tsunami and couldn't be functioned. Mobile phones couldn't be charged in some areas owing to power blackout. Those situations made bad condition for collecting the location data. Sekimoto et al.[2)] also mentioned that the amount of collection of auto GPS data in Tokyo Metropolitan area decreased significantly on the next day of the earthquake. Therefore, we need to understand the earthquake effect on data lacking.

The mesh-unit population calculated from the data is shown in Figure 1. The value of populations are summed up daily for two months, from 1st March to 30th April 2011, and the maximum, the minimum and the average value are plotted. The minimum values of 12th and 13th March are over 100 thousand less than before the earthquake. This can be thought of as the bad effect of the deteriorated condition for collecting data; consequently, the two days are excluded from the research period.
The values of population after 14th March are still unstable. Therefore, the rate of population will be used instead of the absolute number of population as mentioned in section 2.2.

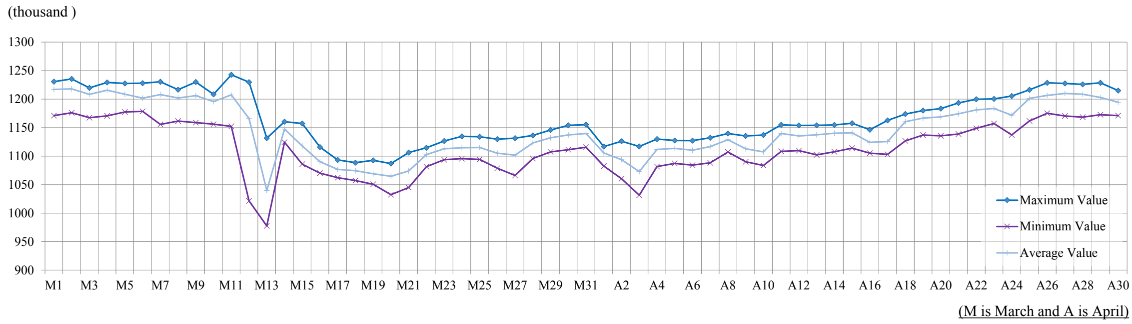

Figure 1. Daily mesh-unit population in Sendai city

3. The characteristic of population rate distribution in normal time

Change of the population rate is used to understand mobility. And to analyze the influence of disaster on mobility decline, the research term is divided into before and after the earthquake. 7 days from 4th to 10th March 2011 are defined as the before term named 'normal time' in this study.

To understand the Sendai city mobility in normal time, the distributions of population rate ($P_{d,t}$) in the before term are classified. Correlation coefficients between every distribution of population rates are computed and the correlation coefficients are used for cluster analysis (group average method). According to the results of the analysis, the distributions of population rate were classified into 4 clusters shown in Table 1.

Table 1. Classified cluster of time interval of population rate distribution

Group A	2:00, 3:00, 4:00, 5:00, 6:00, 7:00
Group B	8:00, 9:00, 10:00, 11:00, 12:00, 13:00, 14:00, 15:00, 16:00, 17:00
Group C	18:00, 19:00, 20:00, 21:00
Group D	none corresponding time interval

* 2:00 means from 2:00 to 3:00

** All the above time intervals are categorized weekdays' only. Thus, weekends are excluded from normal time interval from here.

Here, we define that distribution of population rate in weekdays the during the normal time interval *t* ($Q_t = \{q_t^1, \cdots, q_t^M\}$) is [normal time t:00]. The information of Q_t is displayed on the Sendai map, and the characteristics of population rate distribution in normal time interval are elucidated as Table 2.

Table 2. The characteristics of population rate distribution in normal time interval

[normal time 2:00] : Q_2	High rates of population are distributed in suburbs of Sendai city and in nightlife district called Kokubun-cho Town.
[normal time 9:00] : Q_9	High rates of population are distributed near Sendai station, business district, and government office quarter as the result of commuters.
[normal time 14:00] : Q_{14}	High rates of population are distributed over a wide area of downtown on account of not only employees and students but also shoppers.
[normal time 18:00] : Q_{18}	High rates of population are distributed not only Sendai station but also commercial areas and arcaded street because of shopping after work.

4. Analysis of mobility decline and restoration in disaster

4.1. Determination of time interval set of explanatory variable

It can be considered that the distribution of population rate after the earthquake ($P_{d,t}$) is different from it before the earthquake (Q_t) because of the mobility decline in disaster. But the Q_t may restore similarity as $P_{d,t}$ with the lapse of time. It is called the mobility restoration and the restoration process of mobility will be analyzed using a multiple regression analysis.

The population rates after disaster ($P_{d,t} = \{p^1_{d,t}, \cdots, p^M_{d,t}\}$) are expressed with the three explanatory variables from $Q_t = \{q^1_t, \cdots, q^M_t\}$ based on Least Squares Minimization as the following equation (2). And the three coefficients, ($\alpha_{a,(d,t)}, \alpha_{b,(d,t)}, \alpha_{c,(d,t)}$) are estimated using a multiple regression analysis. The coefficients, $\alpha_{a,(d,t)}, \alpha_{b,(d,t)}, \alpha_{c,(d,t)}$ are to be the contribution rate of explanatory variables because of the constraint condition.

$$min. \sum_{m\in\{1,\cdots,M\}} \left\{p^m_{d,t} - \left(\alpha_{a,(d,t)} q^m_{T\in A} + \alpha_{b,(d,t)} q^m_{T\in B} + \alpha_{c,(d,t)} q^m_{T\in C}\right)\right\}^2 \quad (2)$$

$$subject\ to$$

$$\alpha_{a,(d,t)} + \alpha_{b,(d,t)} + \alpha_{c,(d,t)} = 1$$

$$\alpha_{a,(d,t)}, \alpha_{b,(d,t)}, \alpha_{c,(d,t)} \geq 0$$

Explanatory variables should be satisfied the following 3 conditions;

1) Explanatory variables are selected one in each classified clusters of Table 1 (Group A, B and C) to solve multicollinearity.
2) One of the time intervals of explanatory variable should be included in the same time interval of objective variable to quantify the mobility restoration.
3) One time interval of objective variable is expressed with the same three time interval of explanatory variable in all days.

The results of regression analysis are shown in Table 3. The target period is weekdays from 14th March to 30th April (33 days) after the disaster. A time interval of objective variable consists of three time intervals of explanatory variable and fulfilled the 1st condition. The bold time intervals in each set are represented satisfaction of the 2nd condition. And three time intervals of explanatory variable are same regardless of the date as the 3rd condition.

Table 3. Sets of time interval of explanatory variables (Q_t)
in each time interval of objective variable ($P_{d,t}$)

t for Q_t / t for $P_{d,t}$	Group A	Group B	Group C
2:00	**[normal time 2:00]**	[normal time 8:00]	[normal time 21:00]
3:00	**[normal time 3:00]**	[normal time 8:00]	[normal time 21:00]
4:00	**[normal time 4:00]**	[normal time 9:00]	[normal time 21:00]
5:00	**[normal time 5:00]**	[normal time 9:00]	[normal time 21:00]
6:00	**[normal time 6:00]**	[normal time 9:00]	[normal time 18:00]
7:00	**[normal time 7:00]**	[normal time 9:00]	[normal time 18:00]
8:00	[normal time 2:00]	**[normal time 8:00]**	[normal time 18:00]
9:00	[normal time 2:00]	**[normal time 9:00]**	[normal time 18:00]
10:00	[normal time 2:00]	**[normal time 10:00]**	[normal time 18:00]
11:00	[normal time 2:00]	**[normal time 11:00]**	[normal time 18:00]

12:00	[normal time 2:00]	**[normal time 12:00]**	[normal time 18:00]
13:00	[normal time 2:00]	**[normal time 13:00]**	[normal time 18:00]
14:00	[normal time 2:00]	**[normal time 14:00]**	[normal time 18:00]
15:00	[normal time 2:00]	**[normal time 15:00]**	[normal time 18:00]
16:00	[normal time 3:00]	**[normal time 16:00]**	[normal time 18:00]
17:00	[normal time 5:00]	**[normal time 17:00]**	[normal time 19:00]
18:00	[normal time 2:00]	[normal time 9:00]	**[normal time 18:00]**
19:00	[normal time 2:00]	[normal time 9:00]	**[normal time 19:00]**
20:00	[normal time 2:00]	[normal time 9:00]	**[normal time 20:00]**
21:00	[normal time 2:00]	[normal time 9:00]	**[normal time 21:00]**

4.2. Quantitative analysis of mobility changes

In this section, the distribution of mobility after the earthquake is described with the distributions of mobility in normal time according to the results of section 4.2.

Figure 2 and Figure 3 show coefficient value of each explanatory variable when the time interval of objective variable is 14:00 and 18:00 respectively. The values of coefficient from 1st to 10th March are also plotted in the same graph to compare between before and after disaster. The coefficient values of explanatory variable represent the distribution of mobility because they are contribution rate of explanatory variables.

Case 1) The time interval of objective variable is 14:00 (Figure 2)

The coefficient value of [normal time 14:00] is less than 0.5 but [normal time 2:00] is more than 0.5 just after the earthquake (from 14th to 16 March). This means that about 50% of people couldn’t take the normal time mobility and might stay at home. But the coefficient value of [normal time 14:00] is getting increase with the lapse of time and reaches more than 0.8 from 24th March. That means about 80-90% mobility was restored from that point.

The coefficient value of [normal time 18:00] rises about 0.1-0.2 from April. It might be that people started to go shopping or go to downtown but this mobility was still different from normal time.

There was a big aftershock at midnight of 7th April. The effect of it appeared on 8th April. The distribution of [normal time 18:00] had been stopped and [normal time 2:00] increased remarkably.

Case 2) The time interval of objective variable is 18:00 (Figure 3)

Almost 100% people couldn’t secure [normal time 18:00] distribution from 14th to 18th March according to Figure 3. And about 60% people were just at home (like 2:00) and 40% took commuting behavior (like 9:00) at 18:00 for 4 days (until 17th March). This [normal time 9:00] distribution has been switched to [normal time 18:00] little by little every day, but distribution of [normal time 18:00] is hardly restored until 22nd April.

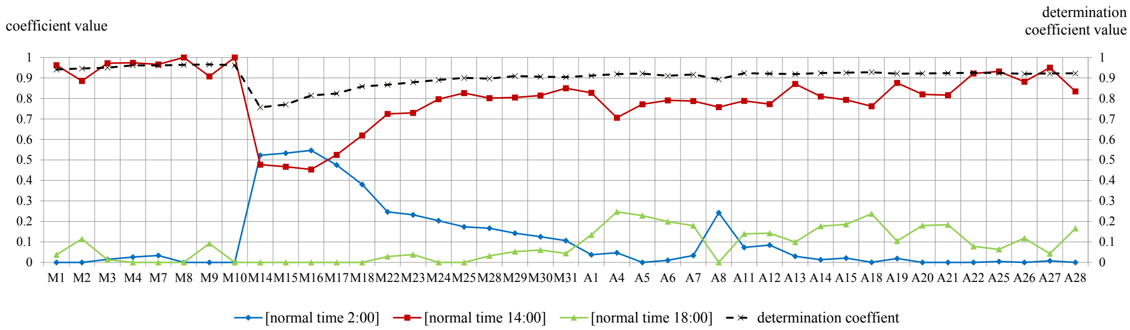

Figure 2. Coefficient value when the time interval of objective variable is 14:00

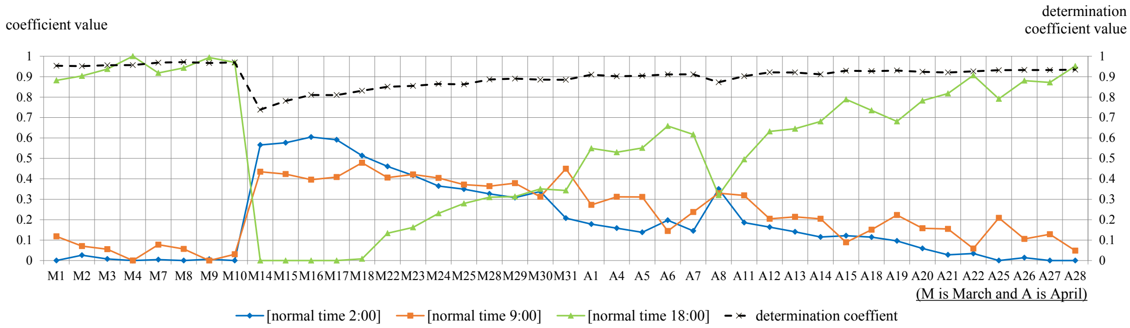

Figure 3. Coefficient value when the time interval of objective variable is 18:00

5. Conclusions

We suppose the population distribution is consequences of mobility in this study. The population distribution could be obtained from GPS positioning data but they couldn't be used as it is because of missing data after the earthquake. Therefore, the rate of population was used instead of population. And we could confirm human mobility decline and restoration process after the Great East Japan Earthquake according to the analysis of the change of distribution of population rate before and after the earthquake.

The change of the rate of population after the earthquake could be expressed with the rate of population before the earthquake using the computed results of multiple regression analysis. And some of the results obtained were as follows;

[1] For example, we analyzed when the time is 14:00 -15:00. Just after the earthquake, only 50% people could keep normal time mobility, and the rest of people stayed at home like 2:00 pattern. About 80-90% mobility was restored from 24th March, whereas the rest was still different from the normal time.

[2] The pattern of time interval in 18:00-19:00 was different as mentioned above. About 60% people were at home (like 2:00) and 40% took commuting behavior (like 9:00) at 18:00 for 4 days (until 17th March). The 18:00 mobility pattern was hardly restored until 22nd April.

We can estimate the rate of population after disaster by using the normal time population. This study will be useful to project the public facilities while taking the mobility changes into account, not only for service provision in normal situation but also for disaster reduction after disaster occurred.

This this study, the mobility decline and the restoration process were analyzed using quantification method. However, a factor of restoration (e. g. repaired railway, gasoline shortages resolution, etc.) didn't be mentioned. The relationship between the restoration process and the factor should be specified.

References

1) Takashi Akamatsu, Minoru Osawa, Takeshi Nagae, Hiromichi Yamaguchi, Spatio-Temporal Analysis of Gasoline Shortages in the Tohoku Region after the Great East Japan Earthquake, Special Issue (Invited Papers) on 2011 Great East Japan Earthquake, Journal of JSCE Vol.1, No.1, pp.447-469, 2013
2) Yoshihide Sekimoto, Toshikazu Nakamura, Yusuke Masuda, Hiroshi Kanasugi, Human mobility analysis in the Great East Japan Earthquake in Tokyo Metropolitan area using large-scale GPS data, Proceeding of infrastructure planning, Vol. 45, CD-ROM, 2012