Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 11 (2015) 399 – 412

Transportation Research Procedia

www.elsevier.com/locate/procedia

10th International Conference on Transport Survey Methods

Behaviour analysis using tweet data and geo-tag data in a natural disaster

Yusuke Hara[a]*

[a]*Graduate School of Information Sciences, Tohoku University, 6-6-06, Aoba, Aramaki, Aoba-ku, Sendai, Japan*

Abstract

This paper clarifies the factors that resulted in commuters being unable to return home and commuters' returning-home decision-making process at the time of the Great East Japan Earthquake using Twitter data. First, to extract the behavioural data from the tweet data, we identify each user's returning-home behaviour using support vector machines. Second, we create nonverbal explanatory factors using geo-tag data and verbal explanatory factors using tweet data. Following this, we model users' returning-home decision-making using a discrete choice model and clarify the factors quantitatively. Finally, we show the usefulness and the challenges of social media data for travel behaviour analysis.

Peer-review under responsibility of International Steering Committee for Transport Survey Conferences ISCTSC

Keywords: travel behaviour analysis in a disaster; returning-home behaviour in a disaster; information extraction from social media data

1. Introduction

The 2011 earthquake off the Pacific coast of Tohoku, often referred to in Japan as the Great East Japan Earthquake, was a magnitude 9.0 undersea megathrust earthquake that occurred at 14:46 Japan Standard Time on March 11, 2011. The focal region of this earthquake was widespread, spanning approximately 500 km from north to south (reaching from off the Ibaraki shore to the Iwate shore) and approximately 200 km from east to west. The number of deaths and missing persons attributed to this disaster totalled more than 19,000, and the complex, large-scale disasters of an earthquake, tsunami, and nuclear power plant accident had a major impact on people's lives. The strong earthquake also hit the Tokyo metropolitan area, where it resulted in various traffic problems; for

* Corresponding author. Tel.:+81-22-795-7497; fax:+81-22-795-7494.
*E-mail address:*hara@plan.civil.tohoku.ac.jp

Peer-review under responsibility of International Steering Committee for Transport Survey Conferences ISCTSC
doi:10.1016/j.trpro.2015.12.033

example, many railway and subway services suspended their operations to scan for the potential damage produced by the earthquake. Consequently, virtually every railway and subway user was unable to return home easily; they were called “victims unable to return home”. According to the Measures Council (2012), the number of victims unable to return home that day because of the disruption of transport networks was approximately 5.15 million, 30% of which were people leaving the city that day.

The problem of victims unable to return home in the Tokyo metropolitan area is extremely important for preparing for the next disaster. Although questionnaires were completed after the event, what influenced the returning-home decision-making process after the earthquake disaster has not yet been shown clearly. In addition, great confusion occurred at the time of the disaster, causing victims to forget the details of their location and mental situation. However, the raw information of human behaviour at the time of the disaster is essential information for analysing the evacuation and return-home behaviour.

Some previous studies have examined human behaviour through analysis of behaviour log data at the time of large-scale disasters. Because no rapid and accurate method existed to track population movements after the 2010 earthquake in Haiti, Bengtsson et al. (2011) used position data from subscriber identity module (SIM) cards from the largest mobile phone company in Haiti to estimate the magnitude and trends of population movements after this earthquake and the subsequent cholera outbreak. Their results indicated that estimates of population movements during disasters and outbreaks can be acquired rapidly and with potentially high validity in areas of high mobile phone usage. Lu et al. (2012) also used the same data in Haiti to determine that 19 days after the earthquake, population movements caused the population of the capital, Port-au-Prince, to decrease by approximately 23%, and that the destinations of people who left the capital during the first three weeks after the earthquake were highly correlated with their mobility patterns during normal times, specifically, with the locations of people with whom they had significant social bonds. Lu et al. (2012) concluded that population movements during disasters may be significantly more predictable than previously thought. Overall, these previous studies clarified human movements over long periods of time; they showed that people in areas affected by an earthquake take refuge temporarily and that the population in the affected area recovers over several months. Behaviour log data should be able to clarify not only such long-term human behaviour but also human behaviour at the time of a disaster itself.

In this paper, we analyse tweet data from Twitter as the behaviour log data at the time of the Great East Japan Earthquake. There is much literature on using secondary data such as social media data for monitoring and understanding some events. These studies are called “social sensor” research because people using social media generate information on target events such as physical sensors. Sakaki et al. (2010) considered spatiotemporal Kalman filtering, which is similar to space-time burst detection, to track the geographical trajectory of hot spots of tweets related to earthquakes. Signorini et al. (2011) and Louis and Zorlu (2012) showed expanding disease outbreaks by Twitter data. Majid et al. (2013) indicated travellers’ preferences from online photo-sharing sites such as Flickr. Shelton et al. (2014) used Twitter data related to Hurricane Sandy to uncover broad spatial patterns within this data and showed how these data reflect the lived experiences of the people creating the data.

The amount of research that aims to monitor traffic using social media is increasing. Traffic congestion monitoring can be classified into two categories: one is large-scale traffic monitoring and the other is small-scale traffic monitoring. Most existing large-scale traffic monitoring research has focused on event detection from a large number of social media messages. The research on anomaly detection using social media uses users’ posts as a real-time social sensor. Another approach is a geo-topic model that uncovers the relationship between language distribution and geographical location (Yin et al., 2011; Hong et al., 2012). For small-scale traffic monitoring, Schulz et al. (2013) extracted features from tweets and identified tweets relevant to local and small-scale events. Mai and Hranac (2013) extracted road accidents from Twitter and compared the result with California Highway Patrol traffic incident records. Pan et al. (2013) integrated GPS trajectory data and microblog data to detect anomalous GPS traces. Chen et al. (2014) developed Language-enhanced Hinge Loss Markov Random Fields and indicated the traffic conditions from tweets.

This paper aims to analyse each Twitter user’s travel behaviour, unlike social sensor research that aims to monitor or understand specific events such as the occurrences of earthquakes, disease outbreaks, natural disasters and congestion in traffic networks. Although tweet data do not necessarily contain actual behaviour, there is the possibility they may contain thought processes and behavioural factors. We clarify the factors associated with return-home behaviour in the case of the Great East Japan Earthquake using Twitter data.

2. From tweet data to behaviour data

2.1. Framework

The framework used in our research to analyse users' return-home behaviour using tweet and geo-tag data is shown in Figure 1. The framework comprises the following modules: (1) behaviour inference by tweet data, (2) feature engineering by geo-tag and tweet data, and (3) estimation of the behavioural model. The solid line in Figure 1 shows the data extraction and analysis processes. The dashed line in Figure 1 shows the feature engineering process using other data resources such as road network data and public transport fee data.

In module (1), behavioural inference by tweet data, we infer users' return-home behaviour using support vector machine (SVM) and bag-of-words (BOW) representations. In module (2), feature engineering by geo-tag and tweet data, we take explanatory factors for users' behaviour from tweet and geo-tag data. For instance, the explanatory factors of choice alternatives from geo-tag data are the distance, travel time of each travel mode, and fee. Those factors from tweet data are whether Twitter users checked their family's safety and whether they talked about the reopening of train service. In module (3), estimation of behavioural model, we estimate users' behaviour using a discrete choice model.

Let us show the difference between (1) and (3). In part (1), we preprocess users' tweets and add each Twitter user to the appropriate travel mode category. For example, we add the user who tweeted "I'm very tired because I walked from my office to home for 5 hours" to the category "return home by foot" and the user who tweeted "I will stay at my office overnight because my train has been stopped. Next morning, I will try to return home." to the category "staying in the office or a hotel until the next morning". On the other hand, in part (3), we clarify why some users chose to return home by foot. It is important for policy makers to know whether they returned home by foot because the distance from their office to home was short or because they were concerned about their family.

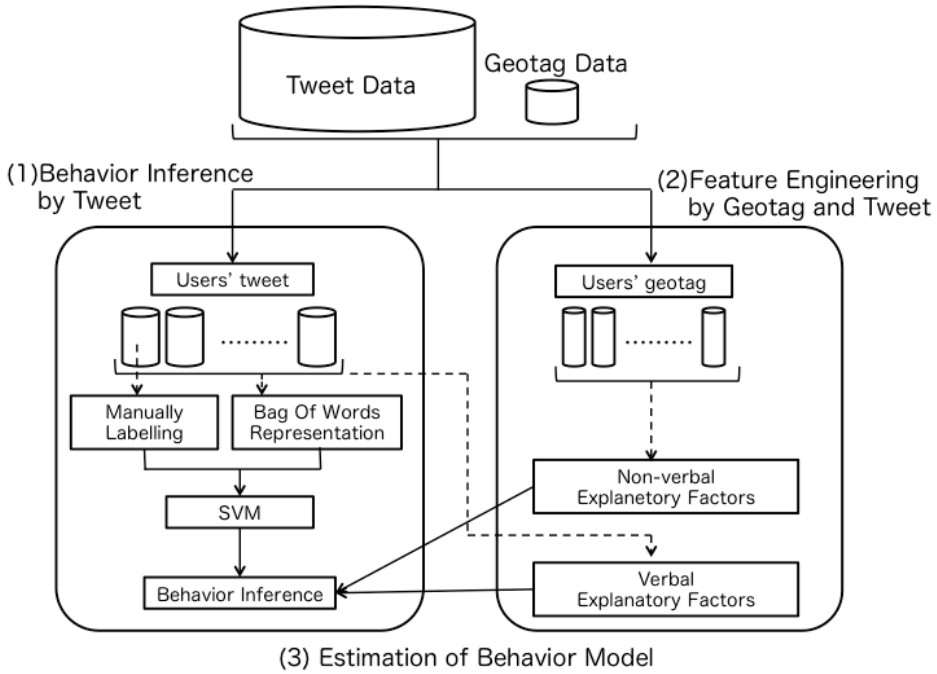

Figure 1. Framework used in our research.

2.2. Data

In this section, we provide an outline of our data. The data comprise approximately 180 million tweets by Japanese people on Twitter from March 11, 2011 to March 18, 2011. In general, Twitter users rarely add their tweets to geo-tag because of the privacy problem. Therefore, there are approximately 280,000 tweets with a geo-tag in the data or 0.1% of all tweets. We extract tweets with timestamps from 14:00 on March 11, 2011 to 10:00 on March 12, 2011 and whose GPS location is within the Tokyo metropolitan area. The number of such tweets is 24,737, and the number of unique users (accounts) is 5,281. To observe users' trips on the day, we extract users with more than two geo-tag tweets, resulting in 3,307 users. We assume that these users could have tweeted about the Great East Japan Earthquake and their return-home behaviour. Consequently, we analyse all tweets from these users from 14:00 on March 11, 2011 to 10:00 on March 12, 2011 (3,307 users, with 132,989 total tweets, 22,763 of which were geo-tagged).

The demographics of social media users differ from those of commuters in the Tokyo metropolitan area in general. Therefore, there is a bias in social media data. To discuss the bias of data from social media, we compare our data with those of other surveys.

It is not easy to label the return-home behaviour of all 3,307 users manually because the number of tweets is 132,989. Reading all tweets and labelling the behaviour of each user requires a very large amount of human resources. Therefore, to solve this problem, this study performed labelling using a support vector machine, and the machine learning technique using small-size supervised data can guess all users' behaviour.

To make supervised data, we tag 300 users' return-home behaviour result manually by reading more than 10,000 tweets. Our label set comprises 1) returning home by foot, 2) returning home by train, 3) staying in the office or a hotel until the next morning, 4) other choice (taxi, bus and others), and 5) unclear. We can identify keywords in these 10,000 tweets to classify the travel mode of each Twitter user.

2.3. Morphological analysis

Next, we conduct morphological analysis using MeCab (2014) and obtain bag-of-words representations of each user's tweets because Japanese sentences do not use separate words as English sentences do. By morphological analysis, the number of unique words is 70,364. These words include words that are important for inferring return-home behaviour and those that are not. Then, we try to find the most important word for our task using supervised data.

We use the information gain to find the relationship between return-home behaviour and each user's tweet. Information gain is an index that shows the decreasing degree of each class's entropy using an existing word, w. If word w is contained in each user's tweet, the random variable X_w equals one; otherwise, $X_w = 0$. The random variable that indicates each class is c, and the entropy, $H(c)$, is written as follows:

$$H(c) = -\sum_{c} P(c) \log P(c). \tag{1}$$

Further, the conditional entropy is written as follows:

$$H(c \mid X_w = 1) = -\sum_{c} P(c \mid X_w = 1) \log P(c \mid X_w = 1),$$

$$H(c \mid X_w = 0) = -\sum_{c} P(c \mid X_w = 0) \log P(c \mid X_w = 0).$$

The information gain, $IG(w)$, of word w is defined as the average decreasing entropy, and is written as follows:

$$IG(w) = H(c) - \left(P(X_w = 1)H(c \mid X_w = 1) + P(X_w = 0)H(c \mid X_w = 0)\right). \tag{2}$$

We calculate all word information gain, $IG(w)$, using five classes: walk, train, stay, other, and unclear. Table 1 shows illustrative examples, and these words have high conditional probabilities in each class. This means that the user tweets the words in each row tending to belong to each class.

Table 1. Illustrative examples of words whose information gain is high.

1) by foot	駅 (station), 歩く (walk), 足 (foot), 休憩 (rest), 自転車 (bicycle), 電車(train), ヤバイ (danger), 止まっ (stop), 半分 (half), 到着 (arrived), 歩ける (can walk), テレビ (TV), トイレ (toilet), 環七 (Kannana Street), km, 川崎 (Kawasaki city), 疲れ (tired), 遠い (far), 道 (road)
2) by train	大江戸線 (O-edo subway line), 入場 (entry), 田園都市線 (Denen-toshi line), 奇跡 (miracle), なんとか (luckily), 順調 (smoothly), 京王線 (Keio line), 乗れる (can take the train)
3) stay	泊め (sleep), 朝(morning), 総武線(Sobu line), 混雑(congested), 検索(search), JR (JR line), 乗車 (take the train), 満員(full capacity), 明け (daylight), 暇 (spare time), 始発 (first train in the morning), 悩む (worry)
4) other	Twitpic
5) unclear	jishin, skype

For example, words that indicate a high probability of walking include “half”, “far”, “km”, “Kawasaki” and “Kannana Street”. They show the user’s location. Further, “toilet”, “tired” and “danger” indicate psychological factors during return-home by foot. It seems curious that the list includes “station” and “train”, but these words were used as “I decided to walk home because the train is stopped” or “the station is very congested because many people wait for reopening train service. I will walk home”.

In the case of train, “miracle” and “luckily” are included, as are “O-edo line” and “Denen-toshi line,” which are the train and subway lines that continued to operate on March 11, 2011. Unlike the walking case, the case of train includes the names of specific train or subway lines.

In the case of staying, “morning”, “daylight” and “sleep” indicate that users slept at a hotel or their office and “first train in the morning”, “worry” and “search” show their return-home timing. Other choices by users such as bicycles and taxis as well as unclear users do not show understandable tendencies. However, they submitted pictures for Twitpic, a photo-sharing site, and tweeted with the #jishin hashtag.

As seen above, the words whose information gain is high are useful for inferring users’ returning-home behaviour. Therefore, we make a classifier using those words as features.

2.4. Support vector machine and behaviour inference

In this section, we infer each user’s behavioural result through support vector machines. We use 300 labelled datasets as supervised data and treat the top 500 unique words of information gain as features of support vector machines. In learning, we perform ninefold cross validation, and the average accuracy rate is 73.3%.

Figure 2 shows the inferred result. The number of users who went by foot was 1,913, the number by train was 359, the number staying was 385, the number of users making other choices was 15, and the number of users whose choice was unclear was 635. This result indicates that the ratio of all returning-home users, with the exception of unclear users, was 84.9%. Therefore, the sample size of this study was 2,672.

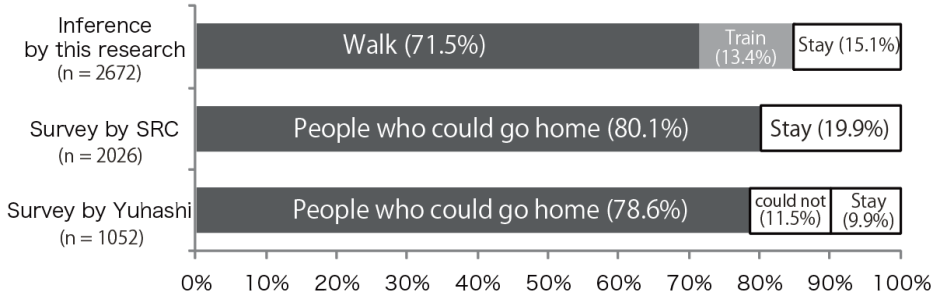

Figure 2. Inferred results and comparison with other surveys.

To verify the accuracy of the inferred results, we compare them with other survey results. Figure 2 shows the results obtained by the Survey Research Center (2011) and Yuhashi (2012). The result from the Survey Research Center is that 80.1% could get home and the result from Yuhashi is that 78.6% could get home. These survey data are obtained by a stratified sampling method (population, gender, age), but our data are raw data from Twitter. In general, young people use Twitter more frequently than older people. Furthermore, these surveys did not ask the transport mode. Although our data and these survey data are different in these aspects, the returning-home inference result from social media data is good enough.

3. Behavioural analysis

3.1. Nonverbal factors

On the basis of the prediction of return-home decision-making classified by user, we created nonverbal and verbal explanation factors from the tweet and geo-tag data and analysed the factors involved in each individual's return-home decision-making.

First, we create the explanation factor about travel behaviour using the geo-tag data classified by the user. In this paper, for simplicity, we assume that the position before the earthquake (14:00 on March 11, 2011) is the location of the office (origin for return-home behaviour) and the position at 10:00 on the day following the earthquake is the home location (destination for return-home behaviour). Next, the road network distance, the time on foot required, the station nearest the office, the station nearest home, the train time required, the train expenses, and the number of times a train change occurs are obtained using GPS data. These are the features created when the network is used normally.

In order to express the spatial spread of people's return-home behaviour, Figures 3a and 3b show the spatial distribution of users' locations before the earthquake and on the day of the earthquake by plotting each user's geo-tag. As an overall trend, the office and home distributions are spatially different, and the home distribution is spread in the direction of the suburban area.

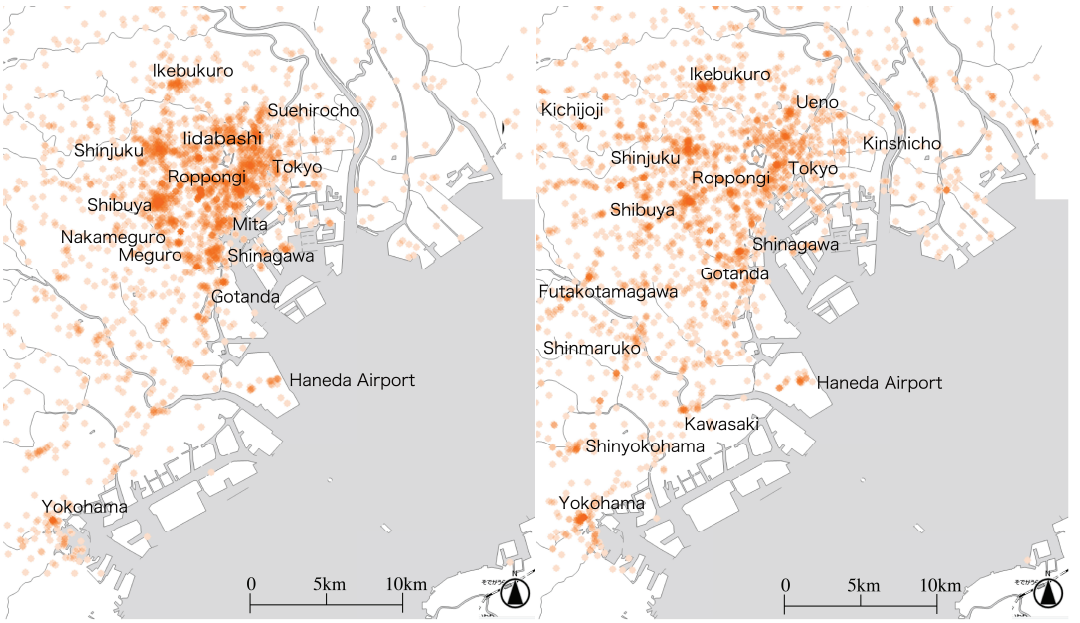

Figure 3. (a) Users' location distribution before the earthquake (14:00, March 11, 2011),
(b) Users' location distribution on the following morning (10:00, March 12, 2011).

Next, the cross-tabulation result of return-home decision-making as a function of the road network distance between office and home is shown in Figure 4. The result indicates that the on-foot rate decreased as distance from home increased, but 50% or more of people went home on foot if their distance was 20 km or more. The survey results by the Survey Research Center (2011) and Hiroi (2011) reported the ratio of people who returned home on foot to all those who returned home. They were 55% when the distance was 20–22 km, 52% (22–24 km), 56% (24–26 km), 34% (26–28 km) and 26% (>30 km). The results inferred from Twitter data have the same tendency as the report.

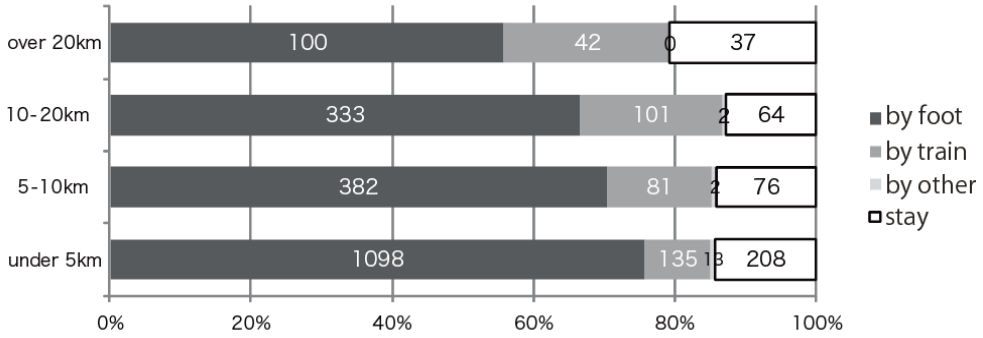

Figure 4. Relationship between return-home behaviour and distance.

Further, Figure 5 shows a timeline of the total trip distance per minute according to different modes of transport at different times of day. This figure is obtained by calculating the distance between each user's geo-tags. The trip distance by foot increased relatively soon after the disaster and the peak of trip distance was achieved around 22:00. The train had not yet resumed at this hour. The trip distance of train users increased from 22:00, and the peak time was at 23:30. The trip distance for staying users increased from 7:00 to 10:00 the following day. These results agree with the inference results by support vector machines. The number of geo-tag tweets per user is approximately 7.4, and this number seems small to understand travel behaviour. However, Figure 5 indicates the travel behaviour patterns by transport mode on the earthquake day in detail.

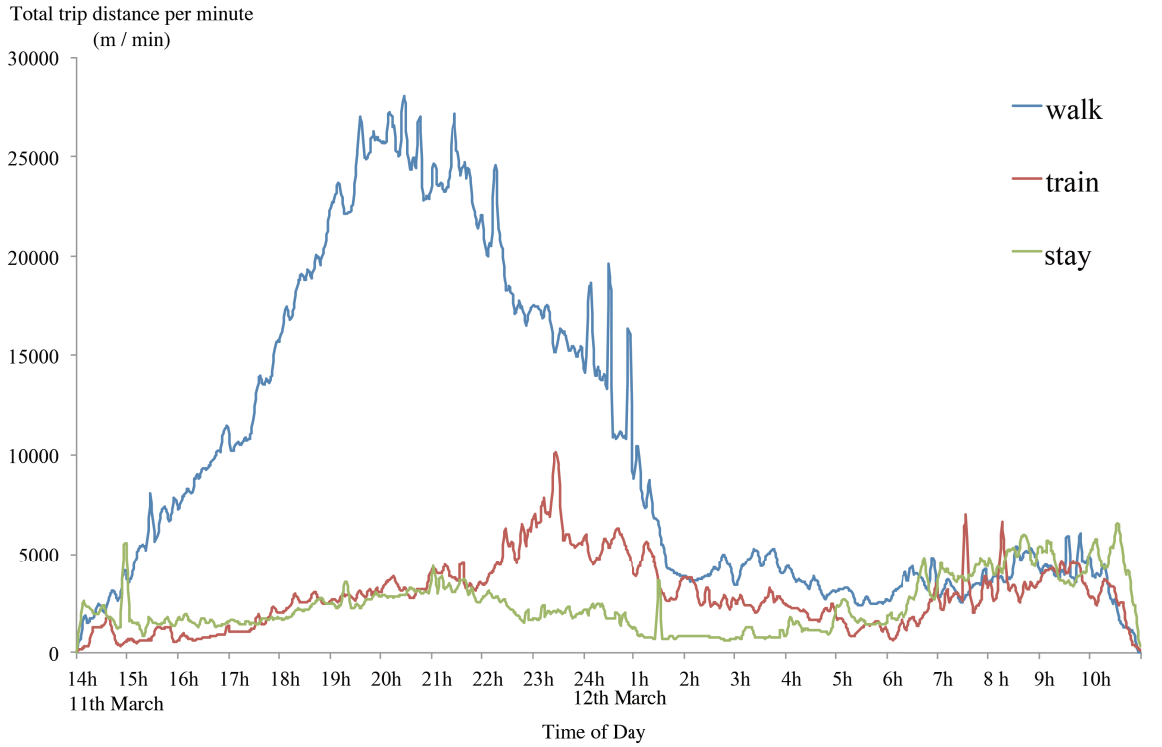

Figure 5. Timeline of the total trip distance per minutes for different transport modes.

3.2. Verbal factors

We generate the verbal explanatory factors from tweet data. We hypothesise that return-home decision-making was affected by not only physical factors (the distance between office and home, travel time and others) but also environment and mental condition. Therefore, we need to extract these explanatory factors from each user's tweets. However, looking at return-home results, tweet contents can include a self-selection bias. The average numbers of tweets for users returning home by foot, by train, and by other modes and those staying in the city are 22.6, 51.8, 80.2 and 61.4, respectively. Train users and those who stayed in the city overnight tweeted more than twice as much as the users who walked home. As train users and those who stayed in the city overnight had more time to use Twitter, this result fits with intuition. Many tweets can include many topics; therefore, the ratio for total tweets is more important than the frequency.

First, we analyse the effect of a safety check with family. In this paper, a family is defined as a spouse and children living together. In total, 353 of the 3,307 subjects spoke about the existence of a family living together. We extract safety check tweets such as "I got an e-mail from my wife! I'm relieved," "I was finally able to contact my wife and daughter by telephone!" and "I could not get through to my son's nursery school by telephone". Figures 6a and 6b show the rates of the safety check tweets and the safety unidentified tweets at different times according to return-home decision-making. Safety check tweets are concentrated before 18:00 (42% for those on foot, 45% for those by train, and 65% for those staying at their office). Safety unidentified tweets are also concentrated before 18:00. We assume that the safety unidentified tweets strongly reflect each individual's psychological state because they remain at every time until safety is checked. If we assume that the tweets at earlier times are more important for each user, a foot-returning user would have regarded the unknown state of their family's safety as more questionable than a train-returning user, and this might have prompted the user to make the decision to return home on foot.

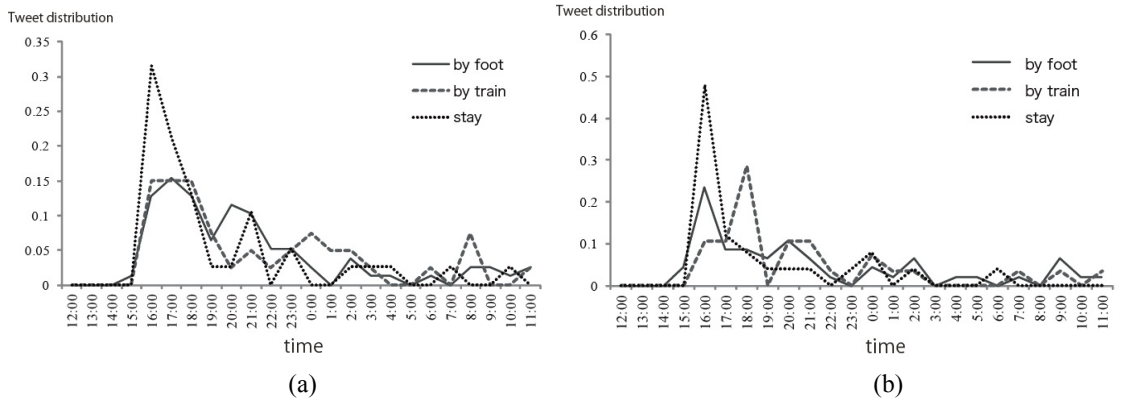

Figure 6. (a) Distribution of safety check tweets, (b) Distribution of safety unidentified tweets.

Next, we analyse the relationship between the information about the trains recommencing operation and returning-home decision-making. Some train services restarted their operations after 20:40 on March 11. Return-home decision-making may have depended on the acquisition of the train-recommencement information. As we cannot observe whether each user could obtain this information, we use tweets about trains restarting. Examples of such tweets include "Ginza subway line is now restarting!" and "It's a miracle! My Keio-line is restarting operation. I can return home!". Figure 7 shows the relationship between the rate of train-restarting tweets and return-home decision-making. The figure indicates that people who chose the train tended to speak about train restarting information. This result does not necessarily indicate a causal relationship between train restarting information and people choosing to return home by train; however, there is a clear difference in the return-home choice between users who tweeted about train reopening and users who did not tweet.

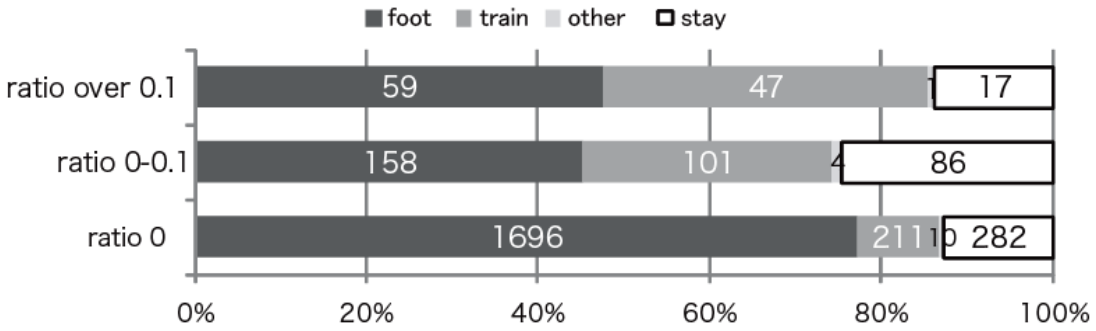

Figure 7. Relationship between train-reopening tweets and return-home decision-making.

Finally, we analyse the relationship between individual psychological factors and return-home decision-making. On March 11, many people talked about their mental situation, in particular, their feelings of fear and anxiety. For example, there were many tweets such as "The earthquake is scary. I don't want to be alone overnight" and "Aftershocks of the earthquake are occurring very frequently. I'm anxious". This psychological factor is different from users' concerns about their families, and we call tweets that include the psychological factor "uneasy tweets". We label uneasy tweets manually by reading them. Figure 8 shows the ratio of uneasy tweets by the return-home decision-making result. Interestingly, people whose rate of uneasy tweeting was under 5% tended to stay at the office or a hotel, whereas people whose utterance rate of unease was over 5% tended to return home on foot. This result shows that people who felt slightly uneasy tended to stay at the office overnight; on the other hand, people with a great anxiety tended to walk home.

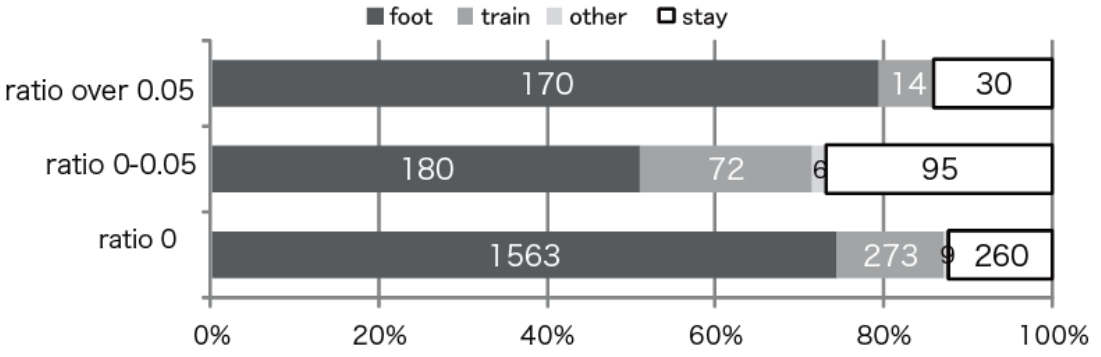

Figure 8. Relationship between uneasy tweets and return-home decision-making.

4. Behavioural model

4.1. Discrete choice model

We build a discrete choice model on the basis of the explanatory variables generated in Section 3. A discrete choice model, also called a random utility model, is a statistical model used in fields such as econometrics, travel behaviour analysis, and marketing (Ben-Akiva and Lerman, 1985; Train, 2003). In this paper, the multinomial logic model (MNL), the most fundamental discrete choice model, is used.

Discrete choice models describe decision makers' choices among alternatives. A decision maker, labeled *n*, faces a choice among *J* alternatives. The decision maker obtains a certain level of utility from each alternative. The utility that decision maker *n* obtains from alternative *j* is $U_{nj} \quad j = 1 \ldots J$. This utility is known to the decision maker but not, as seen below, to the researcher. The decision maker chooses the alternative that provides the greatest utility. The behavioural model is therefore to choose alternative *i* if and only if $U_{ni} > U_{nj} \quad \forall j \neq i$.

Now, consider the researcher. The researcher does not observe the decision maker's utility. The researcher observes some attributes of the alternatives, as faced by the decision maker, labelled $x_{nj} \quad \forall j$, and some attributes of the decision maker itself, labelled s_n, and can specify a function that relates these observed factors to the decision maker's utility. The function is denoted $V_{nj} = V(x_{nj}, s_n) \quad \forall j$ and is often called the representative utility. Usually, *V* depends on parameters that are unknown to the researcher and is therefore estimated statistically.

Since there are aspects of utility that the researcher does not or cannot observe, it is decomposed as $U_{nj} = V_{nj} + \varepsilon_{nj}$, where ε_{nj} captures the factors that affect utility but are not included in V_{nj}. This decomposition is fully general.

The researcher does not know $\varepsilon_{nj} \quad \forall j$ and therefore treats these terms as random. The joint density of the random vector $\varepsilon_n = (\varepsilon_{n1}, \ldots, \varepsilon_{nJ})$ is denoted $f(\varepsilon_n)$. With this density, the researcher can make probabilistic statements about the decision maker's choice. The probability that decision maker *n* chooses alternative *i* is the following:

$$\begin{aligned} P_{ni} &= \Pr(U_{ni} > U_{nj} \forall j \neq i) \\ &= \Pr(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \forall j \neq i) \\ &= \Pr(V_{ni} - V_{nj} > \varepsilon_{nj} - \varepsilon_{ni} \forall j \neq i). \end{aligned} \tag{3}$$

This probability is a cumulative distribution, namely the probability that each random term $\varepsilon_{nj} - \varepsilon_{ni}$ is below the observed quantity $V_{ni} - V_{nj}$. The MNL model is derived under the assumption that the unobserved portion of the utility follows independent and identically distributed (i.i.d.) extreme value distribution:

$$f(\varepsilon_{nj}) = e^{-\varepsilon_{nj}} e^{-e^{-\varepsilon_{nj}}} \tag{4}$$

$$F(\varepsilon_{nj}) = e^{-e^{-\varepsilon_{nj}}} \tag{5}$$

Further, decision maker *n* choosing alternative *i* is derived as follows:

$$P_{ni} = \frac{\exp(V_{ni})}{\sum_j \exp(V_{nj})}. \tag{6}$$

This is the choice probability of the MNL model.

4.2. Setting of the utility function

In the discrete choice model, the observed utility term V_{ni} is generally defined as $V_{ni} = \boldsymbol{\beta}\mathbf{x}_{ni}$. Here, $\boldsymbol{\beta}$ is a coefficient vector and $\mathbf{x}_{ni}$ is an explanatory vector for decision maker *n*'s alternative *i*. In our research, we use datasets comprising the 2,672 samples identified by support vector machines, with the exclusion of persons unclear, and with the set of choices consisting of walking, taking the train, using another mode of transport, or staying at the office. The explanatory variables required for walking are the time by foot, the ratio of uneasy tweets, and an alternative specific constant. The explanatory variables required for train are the travel time by train, the logarithm of the distance between office and home, the ratio of train-restarting tweets, the dummy variables for family safety check tweets, and an alternative specific constant. The explanatory variables for staying are the ratio of uneasy tweets, the ratio of tweets about whether one can stay in the city, the dummy variables for family safety check tweets, and an alternative specific constant. We normalise the utility of the others to zero. The travel time by foot or train is calculated in the normal situation because it is difficult to know the exact travel time during the disaster period. The walking time is almost the same, but the travel time in the period in question is more than twice or three times as long as during the normal situation.

Next, we outline the method used to estimate the coefficient parameter of a utility function. The MNL model's likelihood function is written as follows:

$$LL(\beta) = \sum_{n=1}^{N} \sum_i \delta_{ni} \ln P_{ni} \tag{7}$$

where δ_{ni} is the Kronecker delta; if decision maker *n* chooses *i*, $\delta_{ni} = 1$, otherwise, $\delta_{ni} = 0$. This likelihood function is globally concave, according to McFadden (1974). Therefore, parameters can be uniquely estimated with a maximum likelihood estimation.

4.3. Estimation result

The estimation results for the above settings are shown in Table 2. The likelihood ratio index is 0.428, and its goodness of fit is adequate. Moreover, the coefficient parameter of travel time is negative, which shows that people avoid long trips as the general trend. The coefficient parameter of distance from the office to home is positive, and it indicates that the choice probability of a train increases as the distance increases. These results are suitable for basic analysis and intuition.

Further, we discuss the parameters generated from tweet data. The ratio of train-restarting tweets increases the choice probability of taking the train, and this result matches the basic analysis findings. As it indicates that knowing train restarting information can affect return-home behaviour, it is important for policy makers to disclose information to avoid a chaotic situation. The ratio of uneasy tweets has an interesting tendency. It encourages both returning home by foot and staying at the office. The basic analysis clarifies that users whose ratio of uneasy tweets is low tend to stay at the office and those who tweet uneasily more frequently tend to walk home. The coefficient value indicates the same tendency. The ratio of tweets about whether the user can stay in the city predicts the choice to stay in the city centre. Some people tweeted "My boss commanded us to return home real soon now. So, I'm walking the streets" and "My company tells employees to get out of the office but my home is very far. I don't know what to do until train operation restarts". In contrast, some people tweeted "My boss let us use the office space freely because it is an emergency situation now. He is very kind!". Whether the user can stay in the city centre can affect

the decision-making for returning home. If the ratio of tweets about whether the user can stay in the city centre is high, the probability of staying at the office or in a hotel is high. This indicates that whether users have a waiting position or space in the city centre is the important factor for managing “victims unable to return home”. Finally, the family safety check tweet dummy variable predicts the choice to take the train or to stay in the city centre. These alternatives had the characteristic of postponing the return-home decision-making on that day. If users did not check family safety, they may have chosen to walk home. Therefore, the family safety check is an important factor to avoid confusion in major disasters.

Table 2. Estimation result using the MNL model.

Variables	Estimator	t-value
Travel time (min/10) [foot, train]	−0.012	−2.20
log(distance(km)) [train]	0.36	5.50
Ratio of train-reopening tweet [train]	4.17	5.72
Ratio of uneasy tweet [foot]	6.05	2.71
Ratio of uneasy tweet [stay]	4.52	1.82
Ratio of tweets about whether he/she can stay in the city [stay]	2.98	4.52
Family safety checked tweet dummy [train, stay]	1.14	3.54
Alternative specific constant [foot]	4.88	18.50
Alternative specific constant [train]	2.46	8.48
Alternative specific constant [stay]	3.08	11.61
Observations	2672	
Initial log likelihood	-3704.179	
Final log likelihood	-2107.771	
Likelihood ratio index ($\overline{\rho}^2$)	0.428	

4.4. Sensitivity analysis

Finally, we conduct sensitivity analysis for policy making to avoid confusion in a disaster. One scenario is a situation in which people can stay in the city centre because their employer does not instruct them to return home or because there are many safe shelters in the city centre. Another scenario is the situation in which people can contact their family promptly after a natural disaster using future information and communication technology. Figure 9 shows the results.

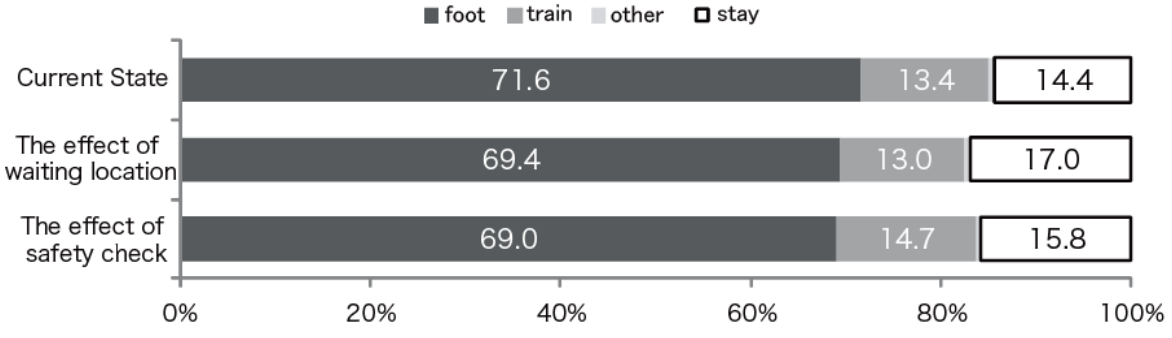

Figure 9. Results of sensitivity analyses.

First, we consider the case where all people have a waiting place. If the ratio of tweets about whether the user can stay in the city to the users who choose foot, train and other options is the same as the average ratio of stay choosers,

the number choosing to stay will increase by 1.18 times and the share of those staying in the city centre will be 17.0%. In contrast, the share of returning-home behaviour on foot and by train would decrease by 3%. Although a 3% reduction seems to be a small influence, in general, traffic congestion and confusion occur in a transport system when more than 10% of the supplied capacity is exceeded. From this point of view, the effect of a 3% reduction is significant for avoiding a chaotic traffic situation.

Next, we analyse the influence of safety checking within a family. From the tweets, we find that there were 353 decision makers with families living together. When all of these 353 individuals were able to check on their family's safety by 17:00, as shown in Figure 8, the number of people who chose the train or to stay increased by 1.1 times, and the number of individuals who chose to go home on foot decreased by a factor of 0.95. Safety checks within a family at the time of a disaster are important information. Because lines of communication other than mobile phones played a major role in this earthquake disaster, these communication tools can help to prevent confusion in the transport network.

5. Conclusions

In this paper, we inferred returning-home behaviour in the Tokyo metropolitan area after the Great East Japan Earthquake using tweet data and geo-tag data from Twitter and clarified the decision-making factors. Although the returning-home behaviour inference method and the behavioural model were based on existing techniques, by combining them with two data sources, the return-home behaviour for each individual and its factors were clarified only on the basis of the Twitter data. Further, we conducted a virtual scenario simulation and analysed the effect of waiting space and communication tools.

This work inferred the return-home behaviour using tweet data and the accuracy of this inference in this research was not bad in comparison with social surveys after the event. Conducting social surveys comes with a high cost to design the survey, collect questionnaire results, and analyse the survey data. On the other hand, Twitter data need analysis only. The factors of return-home behaviour included not only the transport service level and the distance between home and the office but also psychological factors about family concern and information diffusion about the restart of train operation. Twitter data can collect these psychological factors.

Finally, let us show the advantages and disadvantages of our approach. There are three advantages: first, Twitter data are open data. Most social media data are available for everyone and incur no privacy problem. Other secondary data, for example probe data (GPS trajectory data), can come with privacy problems and everyone can collect these data. Second, Twitter data have a strong point for flash report. The accuracy of the analysis of Twitter data may not be very high; however, this approach is useful for discovering latent problems quickly. Third, Twitter data can capture rare events and incidents. Because Twitter is a web service that operates constantly, it is useful to capture rare events such as disasters and big events such as the Olympic games and World Cup. However, there are three disadvantages: the first problem is the representativeness of the data. Social media users are biased, and the population of social media users and the population in social surveys are quite different. The second problem is the ambiguity of classification. Although this study used support vector machines and classified users by their transport mode, the result necessarily included error. Inference by language data only has limitations and requires other data resources to improve its accuracy. The third problem is that of missing data. Social media data may not include what we want to know. In that case, social surveys are useful because they allow such questions to be asked directly.

For future work, this approach needs to clarify the limitations of social media data. In this research, we used tweet data and geo-tag data. Geo-tag data are useful to know accurate locations, but the amount of the data is small because of the privacy problem. However, some tweets can include geographic information such as city name, station name, street name and facility name. Treating these data efficiently is important for ensuring high accuracy.

Acknowledgements

We would like to thank the Great East Japan Earthquake Big Data Workshop and Twitter Japan for their assistance. This work was supported by JSPS KAKENHI Grant Number 25820236.

References

Ben-Akiva, M., Lerman, S., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand, MIT Press, Cambridge, MA.

Bengtsson, L., Lu, X., Thorson, A., Garfield, R., von Schreeb, J., 2011. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti. PLoS Medicine, 8(8), e1001083.

Chen, P.-T., Chen, F., Qian, Z., 2014. Road traffic congestion monitoring in social media with hinge-loss markov random fields. 2014 IEEE International Conference on Data Mining (ICDM), 80–89.

Hiroi, Y., 2011. The report of victims unable to return home in Tokyo metropolitan area under the Great East Japan Earthquake – social survey and analysis – (in Japanese). http://www.cbr.mlit.go.jp/kensei/pdf/reference.pdf

Hong, L., Ahmed, A., Gurumurthy, S., Smola, A.J., Tsioutsiouliklis, K., 2012. Discovering geographical topics in the twitter stream. Proceedings of the 21st International Conference on World Wide Web, 769–778.

Louis, C.S., Zorlu, G., 2012. Can Twitter predict disease outbreaks?. BMJ, 344:e2353. doi: 10.1136/bmj.e2353

Lu, X., Bengtsson, L., Holme, P., 2012. Predictability of population displacement after the 2010 Haiti earthquake. Proceedings of the National Academy of Sciences of the United States of America, 109(29), 11576-11581.

Mai, E., Hranac, R., 2013. Twitter interactions as a data source for transportation incidents. Proceedings of Transportation Research Board 92nd Annual Meeting, 13-1636.

Majid, A., Chen, L., Chen, G., Mirza, H.M., Hussain, I., Woodward, J., 2013. A context-aware personalized travel recommendation system based on geotagged social media data mining. International Journal of Geographical Information Science, 27(4), 662-684.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. Frontiers in Econometrics, Academic Press, New York, 105-142.

Measures Council for victims unable to return home when an earthquake directly hits Tokyo metropolitan area, 2012. Measures council for victims unable to return home by earthquake that directly hits Tokyo Area Final Report (in Japanese). http://www.bousai.go.jp/jishin/syuto/kitaku/pdf/saishu02.pdf.

MeCab, 2014. MeCab: Yet another part-of-speech and morphological analyzer, http://mecab.sourceforge.net/.

Pan, B., Zheng, Y., Wilkie, D., Shahabi, C., 2013. Crowd sensing of traffic anomalies based on human mobility and social media. Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, p344–353.

Sakaki, T., Okazaki, M., Matsuo, Y., 2010. Earthquake shakes Twitter users: real-time event detection by social sensors, Proceedings of the 19th International Conference on World Wide Web (WWW'10), 851-860.

Schulz, A., Ristoski, P., Paulheim, H., 2013. I see a car crash: Real-time detection of small scale incidents in microblogs, In The Semantic Web: ESWC 2013 Satellite Events, 22–33.

Shelton, T., Poorthuis, A., Graham, M., Zook, M., 2014. Mapping the data shadows of Hurricane Sandy: Uncovering the sociospatial dimensions of 'big data'. Geoforum, 52, 167-179.

Signorini, A., Segre, A.M., Polgreen, P.M., 2011. The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic. PLoS ONE 6(5): e19467. doi:10.1371/journal.pone.0019467

Survey Research Center, 2011. Survey of the Great East Japan Earthquake disaster ("victims unable to return home") (in Japanese). http://www.surece.co.jp/src/research/area/pdf/press_19.pdf.

Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge.

Yin, Z., Cao, L., Han, J., Zhai, C., Huang, T., 2011. Geographical topic discovery and comparison. Proceedings of the 20th International Conference on World Wide Web, 247–256.

Yuhashi, H,, 2012. Returning-home situation and information behavior in the Great East Japan Earthquake. Japan Society for Disaster Information Studies 14th Workshop (in Japanese), A-4-2, 140-143.