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.