Tweet Trend Analysis in an Emergency Situation

Takeshi Sakaki
The University of Tokyo
7-3-1, Hongo Bunkyo-ku
Tokyo Japan
sakaki@biz-model.t.u-tokyo.ac.jp

Fujio Toriumi
Nagoya University
Furocho Chikusa-ku
Nagoya Japan
tori@is.nagoya-u.ac.jp

Yutaka Matsuo
The University of Tokyo
7-3-1, Hongo Bunkyo-ku
Tokyo Japan
matsuo@biz-model.t.u-tokyo.ac.jp

ABSTRACT

The Great Eastern Japan Earthquake, which struck Japan on March 11, catastrophically affected all aspects of life: buildings, power plants, human life, etc. Moreover, it caused severe problems related to network infrastructure. We can ascertain the degree of network disorder from network traffic logs. Although we can infer what people did when the earthquake occurred on the Web from network traffic logs, we cannot know it precisely. Social media were used effectively during and after this earthquake, and they left a partial log revealing what people did on the Web during and after the earthquake. Such a log is one of the first logs of people's actions in a time of a catastrophic disaster. As described in this paper, we analyze Twitter logs and attempt to extract what happened in the emergency situation.

1. INTRODUCTION

The Great Eastern Japan Earthquake, which struck Japan on March 11, caused severe damage to many buildings and infrastructure. Network infrastructure was also damaged by this earthquake and many users were unable to connect to the Internet around the time of the earthquake because of network device disorders, power failures, and planned blackouts.

During and immediately after the earthquake, many users attempted to collect information about earthquakes from the Internet and to communicate with friends and relatives through the Internet. Especially social media, including Twitter, were said to be used effectively for information collection about the earthquake and communication among family and friends. Consequently, many user activity logs on Twitter exist in the form of tweets. Therefore it is possible to observe what hap-

pened to web users in relation to the Great Eastern Japan Earthquake by analyzing tweets.

In this paper, we present analysis of social media traffic data and reveal how people's activities on social media were affected by the extremely strong earthquake and the series of events occurring in its aftermath. We attempt to detect what happened to users and network infrastructure by analyzing tweets. Past research [7, 8] has shown that tweets can be used to determine human behavior in other contexts. In this paper, we do the same for the March 11 earthquake.

In section 2, we explain two datasets that we prepared, and describe their preprocessing. In section 3, we analyze an alternation in tweet frequencies in the normal situation and compare it to an alternation in tweet frequencies on March 11. In section 4, we describe analysis of changes of tweet frequencies on an area-by-area basis and describe features of them in several areas. In section 5, we present analysis of changes of tweet frequencies on a medium-by-medium basis. As media to post tweets, we treat media of three types: PCs, feature phones (mobile phones excluding smartphones), and smartphones. Additionally, we compare tweet frequency patterns on a combined basis of area and media. In section 6, we introduce related works. Then, in section 7, we conclude the paper and discuss the use of analysis results for future earthquakes.

In conclusion, several kinds of things occurred around the Great Eastern Japan Earthquake, it can be inferred that some of them affect tweet trends from analysis of the frequency of tweets in our datasets. This paper suggests that it is possible to partially observe network users by analyzing tweet frequencies on Twitter.

2. DATASETS AND PREPROCESSING

We prepared a primary dataset and a supplementary dataset of tweets , as shown in Table 1.

The primary dataset is a set of tweets of many Japanese Twitter users. This dataset is created using the following steps.

1. We crawled all tweets of more than 1.3 million Twitter users, who post Japanese tweets.

Table 1: Datasets.

Dataset name	data source	period
Frequent users	Twitter	Mar. 7 2011 – Mar. 24 2011
Toretter	Twitter	May 1 2010 – Mar. 30 2011

2. We define *frequent Twitter users* as those who post more than a certain number of tweets per day. We choose *frequent Twitter users* from all users we crawled.

3. We extract tweets of *frequent Twitter users* from all tweets we crawled and assemble them as a dataset.

4. We count those tweets by hours and normalize these tweet counts by a tweet count for a specific period of time.

As described in this paper, we crawled more than 300 million tweets posted from March 7 to March 24. To compare amounts of tweets before and after the earthquake on an equal basis, we sampled $218,860$ users who posted more than 5 tweets per day, as *Frequent users*, and extracted 50 million tweets from all tweets we crawled. We used the number of all tweets posted on March 7 for normalization because nothing special happened at that time (no event, such as earthquake, typhoon or big news would have affected users in posting tweets).

The supplementary one is a set of tweets including the keyword "jishin"(*earthquake* in Japanese), as collected by an earthquake detection system *Toretter* during ten months, from May 2010 to February 2011. We prepared this dataset to observe changes of tweets counts in the long term.

3. ANALYSIS OF EFFECTS OF THE EARTHQUAKE ON TWEET FREQUENCIES IN JAPAN

In this section, we analyze effects on tweet frequencies in Japan caused by the Tohoku earthquake by comparing patterns of tweet frequencies before and after the earthquake. First we observe tweet frequency patterns all over Japan in the normal situation by analyzing datasets *Frequent users*
(primary), *Toretter*(supplementary). Second, we observe changes of tweet frequency patterns in Japan after the Tohouku earthquake by analyzing a dataset *Toretter*. We compare tweet frequency pattens between in the normal situation and after the Tohouku earthquake and evaluate effects of the earthquake and scheduled blackouts on tweet frequencies (Scheduled blackouts carried out in Japan in the first week after the earthquake because many power plants had stopped).
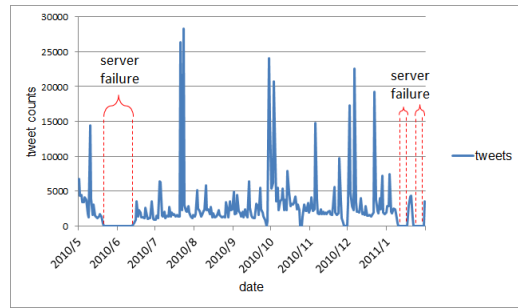

Figure 1: The frequency of tweets including *earthquake* **from May 2010 to February 2011.**
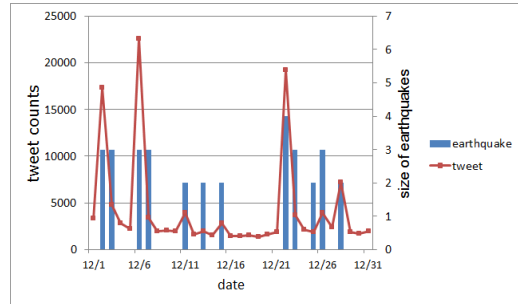

Figure 2: Correlation between tweet frequencies and earthquake occurrences during December 2010.

Judging from these analysis, no severe problems were apparent for most of Twitter users in Japan.

3.1 Observation of tweet frequency patterns in the normal situation

We analyze tweet frequency patterns in the normal situation by observing the frequency of tweets related to earthquake during nine months and daily patterns of tweet frequencies before the earthquake.

Figure 1 depicts changes in the number of tweets of the dataset *Toretter* including keywords "jisin"(*earthquake* in
Japanese) , which were obtained during ten months from May 1 2010 to February 28 2011. This graph shows more tweets than a certain number, about 1500 tweets, excluding some gaps (during June 2010 and January 2011) and some peaks. Those gaps result from our crawler server failure: they can be ignored. It can be inferred that those peaks of the graph are correctly attributable to earthquake occurrences because all tweets include the keyword "jishin". Figure 2 shows changes in the number of tweets including "jishin" and the occurrences of earthquakes in December 2011. We can ascertain the correlation between real earthquakes and tweet counts from Figure 2. The graph shows peaks when earthquakes occurred. Therefore, the amount of tweets in Figure 1 maintained a certain amount excluding effects by earthquakes and server failure.
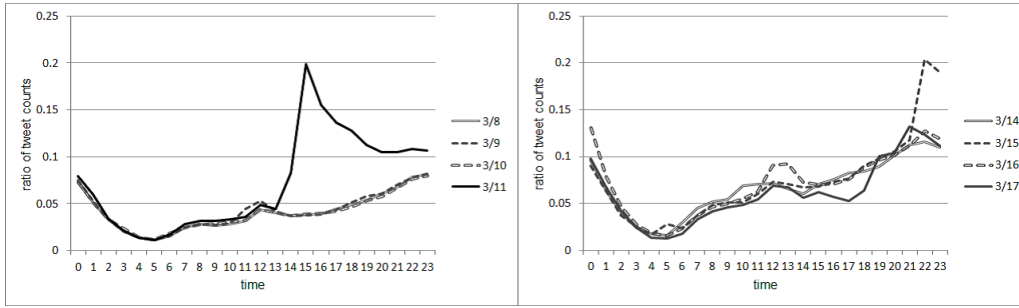
Figure 4: Comparison of daily changes of tweet frequencies on weekdays from March 7 to March 17.
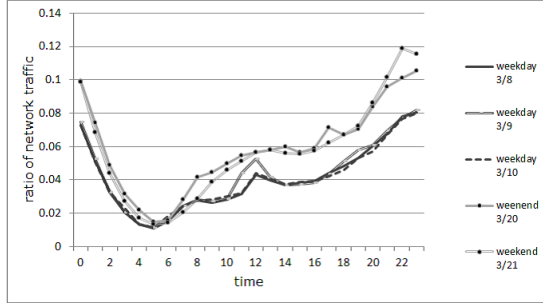

Figure 3: Comparison of weekday and weekend patterns of tweet frequencies in the normal situation.

Figure 5: Correlation between the frequency of tweets and the number of households under blackouts on March 17.

Figure 3 presents data for tweets (*Frequent users*) on weekdays and weekends. In this graph, the frequency of tweets is normalized by the total number of tweets posted at March 7. In Figure 3, graphs of tweet frequencies on weekdays (3/8, 3/9, 3/10) have two peaks. The first peak comes at noon. The second peak comes at night at about 10:00 in the evening. It can be inferred that the first peak comes at noon-time because most companies, factories, and schools in Japan take a lunch break during 12:00–13:00, during which time many people access the Internet. The second peak comes at midnight because many people start to access the Internet after they return home at night.

Similarly, in Figure 3, graphs of tweet frequencies on weekends(3/20, 3/21) show different patterns from those on weekdays. These graphs have a peak at night about 10:00. This feature is different from the feature of patterns in the frequency of tweets on weekdays. It could be because many people do not go to their offices, factories, and schools, or access the Internet and Twitter from their home.

According to these analysis it can be inferred that *Frequent users* post a certain amount of tweets that is steadily flowing and that feature regular patterns.

3.2 Comparison of tweet frequency patterns in Japan before and after March 11
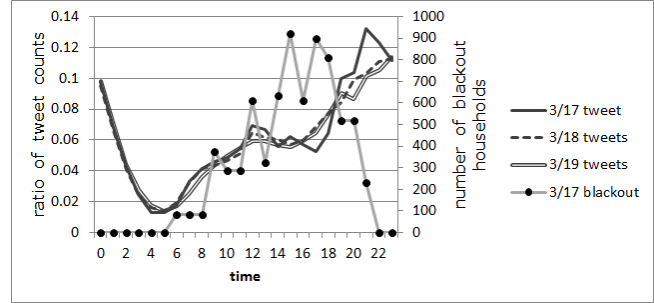
We investigate tweet frequencies in the dataset *Fre-*

quent users for the period around March 11, the day of the Great Eastern Japan Earthquake, and extract anomalous patterns. Figure 4 presents changes in the frequency of tweets that were observed around March 11. In the left graph, the number of tweets increased dramatically at the time of earthquake on March 11, meaning that Twitter was available shortly after the earthquake and that most people posted tweets about the earthquake.

Next, we compare changes in the frequency of tweets around the time of earthquake. Figure 4 also shows the number of tweets on weekdays before the earthquake (drawn in the left side) and after the earthquake (drawn in the right side). In Figure 4, the number of tweets after the earthquake (3/14–3/17) displays a higher value than that before the earthquake(3/8-3/10).

Additionally, the March 17 graph in Figure 4 shows a decline at around 14:00. It might be inferred that this fall resulted from the planned blackout because the largest planned blackout occurred on March 17. Figure 5 expresses changes in the frequency of tweets on Mar. 17, 18 , 19 and the number of households under blackout on March 17. Comparing graphs of Mar. 18 and 19, it appears that the graph on Mar. 17 decreases at around 14:00 and an inverse relation holds between the number of tweets and the number of households under blackout. These drops are not so large.
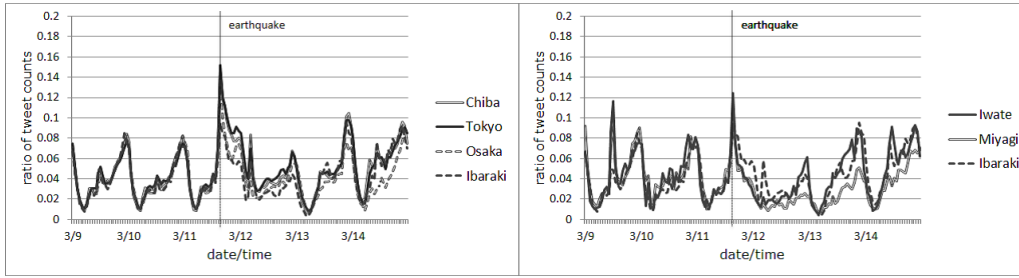
Figure 6: Inter-regional comparison of tweet frequencies from March 9 to March 14.

Table 2: Prefectures treated and their circumstances.

name	damage	blackout	population
Iwate	catastrophic	-	small
Miyagi	catastrophic	-	small
Ibaraki	heavy	-	middle
Chiba	middle	all	large
Tokyo	light	part	very large
Osaka	-	-	large

In fact, no severe problems were apparent for Twitter users in Japan in sending and receiving small text messages such as tweet or mail, judging that most of Twitter users were able to post tweets. However many Japanese Twitter users are concentrated in Tokyo and other large cities. It is difficult to assert that no severe problems using Twitter existed in Japan near the areas affected the earthquake. In the next section, we analyze an alternations in the frequency of tweets on a spatial, not temporal basis.

4. ANALYSIS OF EFFECTS OF THE EARTHQUAKE ON TWEET FREQUENCIES ON A SPATIAL BASIS

In this section, we evaluate effects of the earthquake on tweet frequencies on an area-by-area basis by comparing changes of tweet frequencies in each prefectures in Japan between before and after the earthquake. Twitter user information has a location property. Many users publicize where they live on Twitter. We use this location information of *frequent Twitter users* for tweet analysis.

We analyzed tweet frequencies on a prefectural basis. It is difficult to present all results for Japan's 47 prefectures. Therefore, we select six prefectures as examples as shown in Table 2: Iwate, Miyagi, Ibaraki, Chiba, Tokyo, and Osaka. Miyagi and Iwate are close to the earthquake epicenter. They suffered catastrophic damage from the earthquake and the tsunami. Ibaraki was damaged by earthquakes. In Chiba, many households were cut from the electrical grid because of planned blackouts. In Tokyo, where many earthquakes occurred, many Twitter users live. Osaka is quite distant from the

earthquake center, but it has many Twitter users.

First, we compare changes of tweet frequencies among six prefectures(inter-regional comparison). Next we compare daily changes of tweet frequencies in each prefectures(regional comparison). In all graphs in this section, tweet frequencies are normalized by the total number of tweets posted in each prefectures on March 7.

4.1 Inter-regional comparison of tweet frequencies around March 11

We compare daily changes of tweet frequencies among six prefectures to evaluate effects of the earthquake on tweet frequencies.

Figure 6 depicts the frequency of tweets in these six prefectures from Mar. 9 to Mar. 14. The left graph in Figure 6 depicts the frequency of tweets in less-damaged areas and the right graph depicts tweet frequencies in heavily-damaged areas(the frequency of tweets in Ibaraki is shown in both graphs for comparison).

First, we focus on Mar. 11 in this graph. In that day, a large peak occurred around 15:00. The frequency of tweets increased rapidly after the earthquake, which occurred at 14:46. However, the closer to the earthquake center an area was, the softer the form of the peak in the area is. Additionally, the frequency of tweets is depressed in areas close to the earthquake center: Iwate, Miyagi, and Ibaraki. These two facts show that people who were in areas close to the earthquake center, Miyagi and Ibaraki, posted fewer tweets than people who were distant from the earthquake center. It can be inferred that people who were in areas of a severe earthquake were in too difficult a situation to post tweets and use the Internet, probably because of earthquake damage and escape from the tsunami. Moreover, many network devices were not functional for various reasons at that time.

Second, we focus on before and after the earthquake in Figure 6. The frequency of tweets in Iwate and Miyagi increased on March 9, 10 because small earthquakes occurred as premonitory phenomena. It might be ascertained that people were affected by these tweets. The frequency of tweets in Iwate and Ibaraki decreased on March 11, but returned to the pre-earthquake level during March 13. However, the frequency of tweets
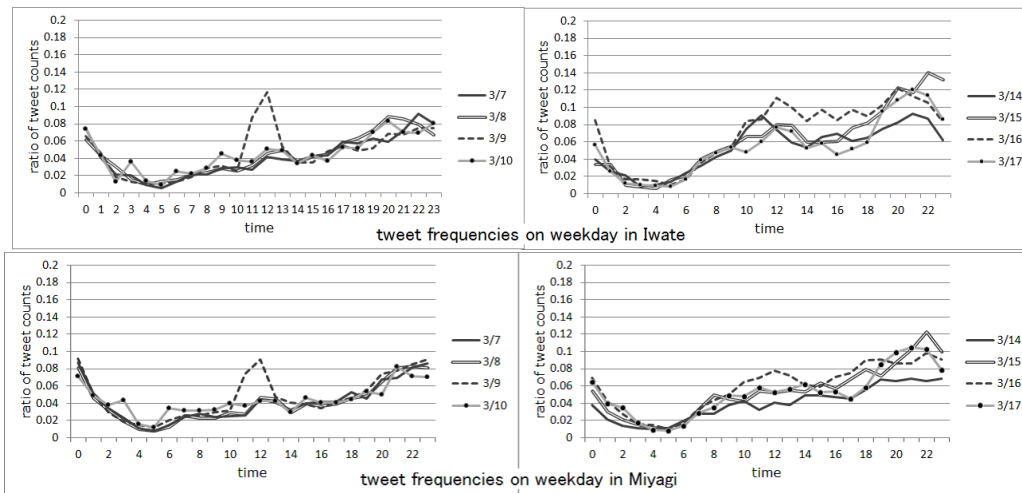
Figure 7: Regional comparison of tweet frequencies on weekdays in heavily-damaged areas(Iwate, Miyagi) .
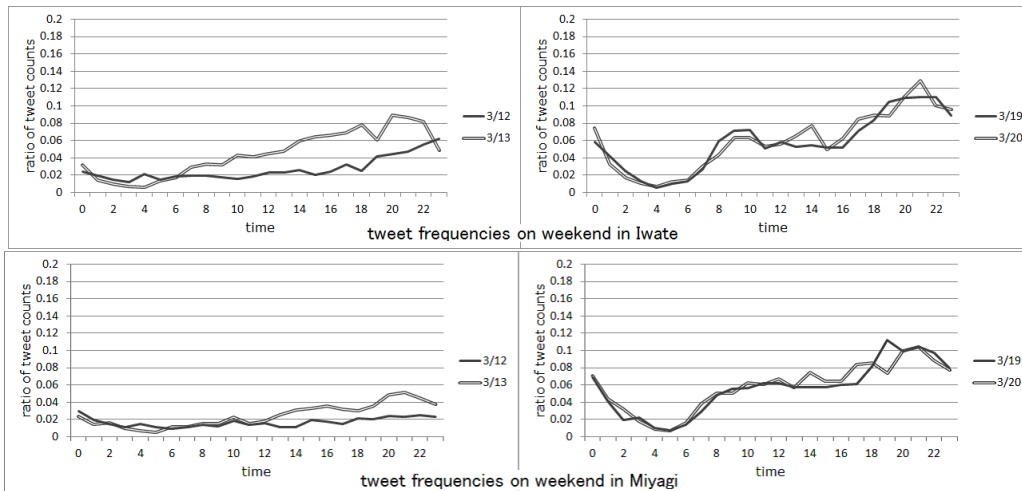

Figure 8: Regional comparison of tweet frequencies on weekends in heavily-damaged areas(Iwate, Miyagi).

in Miyagi dropped on March 11 and did not recover until March 14. It can be said that the situation in Miyagi recovered more slowly than other areas, including Iwate and Ibaraki. Actually, the prefectural capital of Miyagi suffered from the earthquake and the tsunami, unlike other affected areas, such as Iwate and Ibaraki. Additionally, only slight effects on the frequency of tweets was apparent for Osaka and Tokyo. Especially, the amount of tweets increased in Tokyo because aftershocks spur users to post tweets and communicate with others. Therefore the number of tweets increased to a higher level than usual in areas which did not suffer much damage.

4.2 Regional comparison of tweet frequencies around March 11

We compare weekday patterns and weekend patterns

of tweet frequencies around the earthquake in respective areas. Figure 7 depicts the frequency of tweets on weekdays around the earthquake in Iwate, Miyagi. Figure 8 presents the frequency of tweets on weekends around the earthquake in these two areas. Iwate had recovered to normal patterns during 1 week after the earthquake, but Miyagi had not. Results show that it took one week for Iwate to recover Internet access to the normal level, and it took an additional week for Miyagi.

Therefore, people had become able to use the Internet in most areas in a single day, excluding two prefectures: Miyagi and Iwate. It might be inferred that infrastructure for the Internet recovered very quickly for the size of the earthquake. Nevertheless, it is difficult for people in severely damaged areas to use the Internet soon after an earthquake. Therefore, dwindling tweets in one area after earthquakes is likely to indicate that the area was
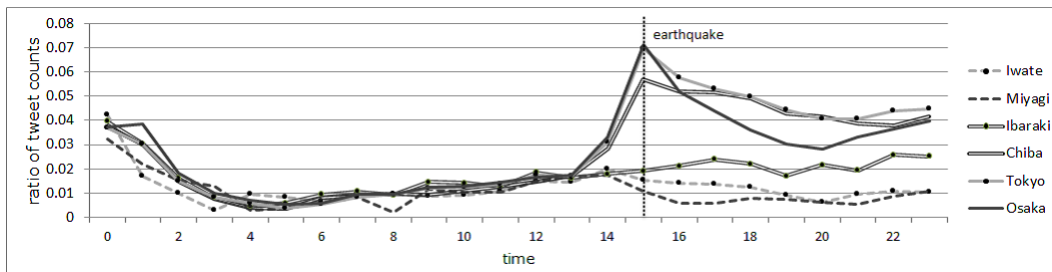
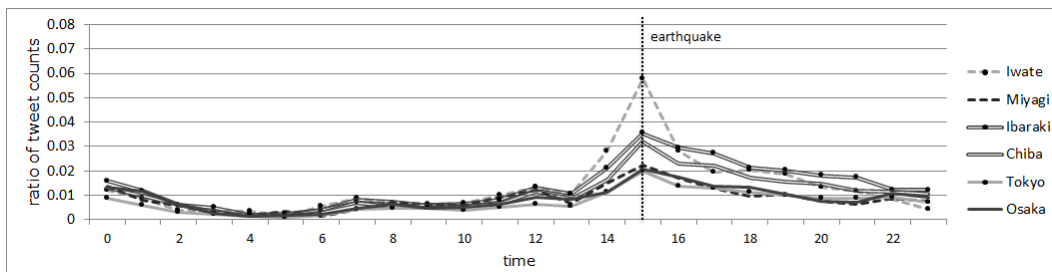Figure 9: Inter-regional comparison of tweet frequencies of PCs on March 11.

Figure 10: Inter-regional comparison of tweet frequencies of feature phones on March 11.
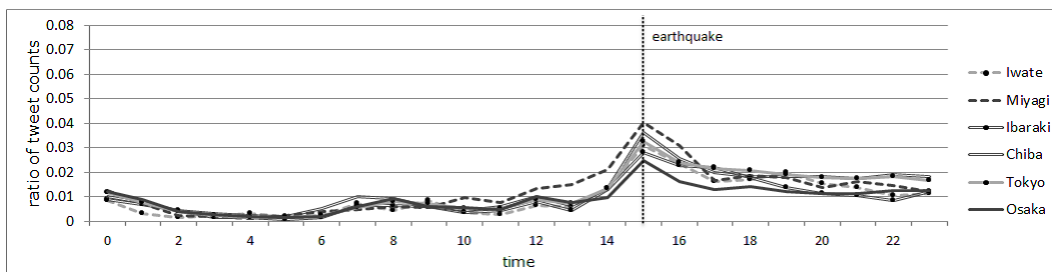

Figure 11: Inter-regional comparison of tweet frequencies of smartphones on March 11.
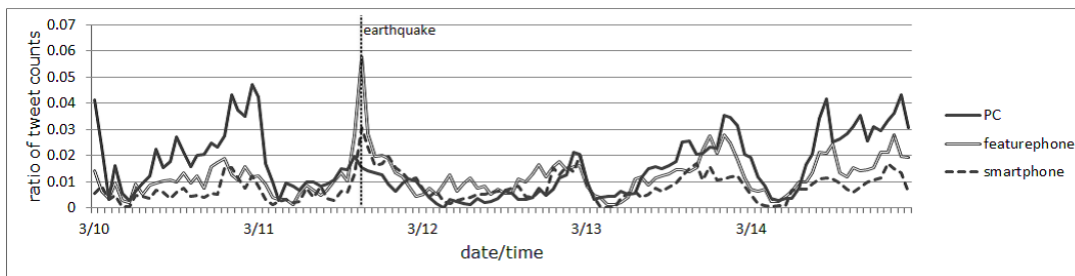

Figure 12: Regional comparison of tweet frequencies of three media in heavily-damaged areas(Iwate) from March 10 to March 14.
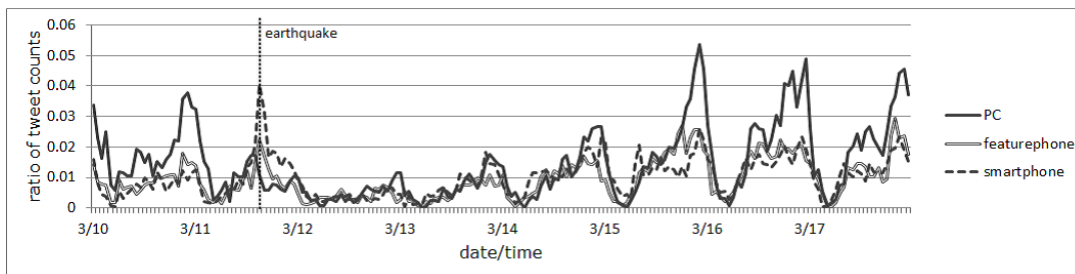

Figure 13: Regional comparison of tweet frequencies of three media in heavily-damaged areas(Miyagi) from March 10 to March 17.

heavily damaged.

5. ANALYSIS OF EFFECTS OF THE EARTH-QUAKE ON TWEET FREQUENCIES ON A MEDIUM BASIS

In this section, we evaluate effects of the earthquake on tweet frequencies on an medium-by-medium basis by comparing changes of tweet frequencies in each media in each prefectures between before and after the earthquake. Each tweet has information about names of tools used to post it. We classify these tools into three groups by medium: PC group Japanese feature phone group, and smartphone group. Japanese feature phones are mobile phones that do not include some functions (application, GPS, etc.) but which can connect to the Internet and send mail.

First, we compare changes of tweet frequencies of three media among six prefectures(inter-regional comparison). Next we compare daily changes of tweet frequencies of three media in these six prefectures(regional comparison). In all graphs in this section , the frequency of tweets is normalized by the total number of tweets posted in each prefectures from each media on March 7.

5.1 Inter-regional comparison of tweet frequencies of three media around March 11

Figures 9,10,11 present tweet frequencies of these three media in these six areas on March 11. Tweets of PCs was dominant in less-damaged areas or undamaged areas: Tokyo, Chiba, and Osaka. The frequency of tweets of feature phones and smartphones were dominant in heavily damaged areas: Iwate, Miyagi and Ibaraki. It was ascertained that feature phones and smartphones functioned and that people were able to connect to the Internet using these devices in an emergency situation. However, PCs did not work at that time for some reasons.

5.2 Regional comparison of tweet frequencies of three media around March 11

We compare changes of tweet frequencies of three media in heavily-damaged areas and less damaged areas.

First, we analyze changes of tweet frequencies of three media in heavily-damaged areas, Iwate and Miyagi. Figures 12,13 present Twitter frequencies of three media in Iwate and Miyagi. Patterns of tweet frequencies in heavily-damaged areas(Iwate and Miyagi) before and after the earthquake differed. After the earthquake, tweet frequencies of PCs in these two areas decreased, apparently. It took several days to rebound to pre-earthquake levels. Additionally, tweet frequencies of feature phones and smartphones decreased a half day after the earthquake. The fact fits our observation that backup batteries of mobile phone base stations are use-
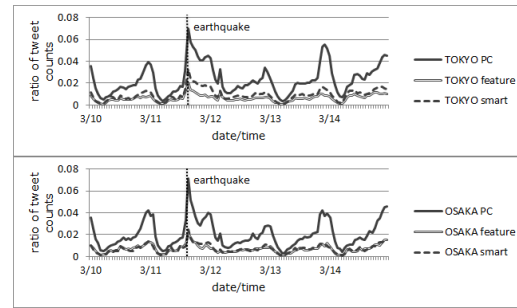

Figure 14: Regional comparison of tweet frequencies of three media in Tokyo and Osaka from March 10 to March 14.
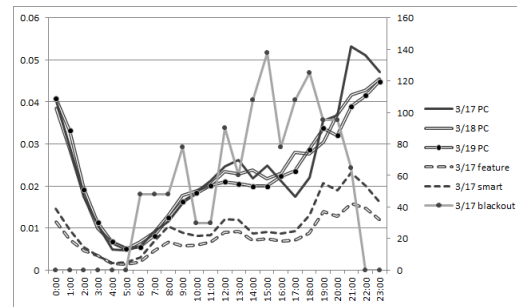

Figure 15: Correlation between tweet frequencies of PC and the number of households under the scheduled blackout in Tokyo March 17

ful for only a half day or one day.

Second, we analyze changes of tweet frequencies of three media in less damaged areas, Osaka and Tokyo. In Figure 14, the frequency of tweets in Osaka and Tokyo after the earthquake was not so different from that before the earthquake. Most tweets were posted by PC. The change in the frequency of tweets in Tokyo on March 12 and 13 closely resembled that before the earthquake. Therefore, these two areas were not damaged by the earthquake, similar to the results of analysis conducted on an area basis.

Figure 15 depicts the frequency of tweets of PC in Tokyo on Mar. 17,18,19 and household counts under planned blackout conditions at that time in Tokyo on March 17. The largest-scale planned blackout occurred around Tokyo and some prefectures on March 17. Certainly, the frequency of tweets of PC on Mar. 17 decreased during the blackout and the frequency of tweets of other media was not different during the blackout, as shown in Figure 15. This fact accords well with the following: it is said that some people were using their mobile phones during the blackout because they were able to use nothing but mobile phones without electricity. Therefore, it might be inferred that planned blackouts affect the the frequency of tweets of PCs.

6. RELATED WORKS

It is known that social media, including Twitter is one of effective tools for crisis communication. There are some researches analyze how people used and utilized Twitter in crisis[4, 2]. Mendoza et al. investigated behaviors of Twitter users in the 2010 Chile earthquake and characterize Twitter in the hours and days following this disaster[5]. Heverin et al analyzed the role of Twitter during violent crimes[3]. Miyabe et al. tried to survey user trend of Twitter after the Great East Japan Earthquake[6]. They analyzed interaction among users on Twitter.

While, some researchers have attempted to observe events that have happened in the real world using social media[1, 8]. For example, Yardi et al. is continuing an investigation of the mode of spreading news and reports that Twitter helps to spread news related to local areas more than news related to broad areas[9]. Sakaki et al. sought to detect earthquakes and to estimate the location of an earthquake epicenter using Twitter[7].

Our research in one such study was undertaken to observe the status of the real world using social media in the face of a crisis. If is different from other researches in that we focus on changes of tweet counts between in the normal situation and in the emergency situation.

7. CONCLUSION

As described in this paper, we concentrate changes in the frequency of tweets and analyze them to observe what happened to network users around the Great Eastern Japan Earthquake.

Our approach revealed the following.

- Tweets has a certain amount and constantly flows in a cyclic pattern (Figure 1, 2, 3).

- Most of Twitter users in Japan did not have critical problems in posting tweets (Figure 4).

- The frequency of tweets was affected by planned blackouts, which were performed several days after the earthquake. (Figure 5, 15).

- The frequency of tweets in heavily damaged areas decreased during several days following the earthquake (Figure 6, 7, 8).

- In heavily damaged areas, tweet frequencies of feature phones and smartphones were dominant shortly after the earthquake, although those of PCs was dominant in less-damaged areas (Figure 9, 10, 11).

- It took a several days for tweet frequencies of PCs to rebound to pre-earthquake levels in heavily-damaged areas. While, tweet frequencies were little affected in less damaged areas with the exception of a few hours after the earthquake (Figure 12, 13, 14).

9. REFERENCES

[1] B. O. Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith. From Tweets to Polls : Linking Text Sentiment to Public Opinion Time Series. *Most*, pages 122–129, 2010.

[2] W. J. Corvey, S. Vieweg, T. Rood, and M. Palmer. Twitter in mass emergency: what NLP techniques can contribute. pages 23–24, jun 2010.

[3] T. Heverin and L. Zach. Microblogging for Crisis Communication: Examination of Twitter Use in Response to a 2009 Violent Crisis in Seattle-Tacoma, Washington Area. In *Proceedings of the 7th International ISCRAM Conference*, Seatle, Washington, 2010.

[4] A. L. Hughes and L. Palen. Twitter adoption and use in mass convergence and emergency events. *International Journal Of Emergency Management*, 6(3/4):248, 2009.

[5] M. Mendoza, B. Poblete, and C. Castillo. Twitter under crisis: can we trust what we RT? In *Proceedings of the First Workshop on Social Media Analytics - SOMA '10*, pages 71–79, New York, New York, USA, jul 2010. ACM Press.

[6] M. Miyabe, E. Aramaki, and A. Miura. Use trend analysis of twitter after the great east japan earthquake. In *Proceedings of SIG-DPS/GN 2011-DPS-148/2011-GN-81/2011-EIP-53*, 2011.

[7] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes Twitter users: real-time event detection by social sensors. *Proceedings of the 19th international conference on World wide web*, pages 851–860, 2010.

[8] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. Predicting Elections with Twitter : What 140 Characters Reveal about Political Sentiment. *Word Journal Of The International Linguistic Association*, pages 178–185, 2010.

[9] S. Yardi and D. Boyd. Tweeting from the Town Square: Measuring Geographic Local Networks. *Proceedings of the International Conference on Weblogs and Social Media*, pages 194–201, 2010.