Information Sharing on Twitter During the 2011 Catastrophic Earthquake

Fujio Toriumi
The University of Tokyo
7-3-1, Hongo, Bunkyo-ku
Tokyo, Japan
tori@sys.t.u-tokyo.ac.jp

Takeshi Sakaki
The University of Tokyo
sakaki@weblab.t.u-tokyo.ac.jp

Kosuke Shinoda
Keio University
kosuke.shinoda@riken.jp

Kazuhiro Kazama
Wakayama University
kazama@sys.wakayama-u.ac.jp

Satoshi Kurihara
Osaka University
kurihara@ist.osaka-u.ac.jp

Itsuki Noda
National Institute of Advanced Industrial Science and Technology
I.Noda@aist.go.jp

ABSTRACT

Such large disasters as earthquakes and hurricanes are very unpredictable. During a disaster, we must collect information to save lives. However, in time disaster, it is difficult to collect information which is useful for ourselves from such traditional mass media as TV and newspapers that contain information for the general public. Social media attract attention for sharing information, especially Twitter, which is a hugely popular social medium that is now being used during disasters. In this paper, we focus on the information sharing behaviors on Twitter during disasters. We collected data before and during the Great East Japan Earthquake and arrived at the following conclusions:
- Many users with little experience with such specific functions as reply and retweet did not continuously use them after the disaster.
- Retweets were well used to share information on Twitter.
- Retweets were used not only for sharing the information provided by general users but used for relaying the information from the mass media.

We conclude that social media users changed their behavior to widely diffuse important information and decreased non-emergency tweets to avoid interrupting critical information.

Categories and Subject Descriptors

J.4 [Computer Applications]: Social and behavioral sciences

Keywords

Twitter, Social networks, Information diffusion, Disaster situation, Earthquake, Information Sharings

1. INTRODUCTION

During a disaster, collecting information is important to save lives. Victims require information about shelters or especially dangerous points. Furthermore, rescuers require information such as victim locations or the availability of supplies. However, in time disaster, it is difficult to collect information which is useful for ourselves from such traditional mass media as TV and newspapers that contain information for the general public. Under the serious disaster situations, mass media cannot send reporters to disaster-stricken areas because traffic and information networks are heavily damaged. Furthermore, victims and rescuer personnel need detailed information that newspersons cannot cover.

Social media attract attention for their information sharing capabilities, especially Twitter, which is one hugely popular social medium that is used during disasters [6][9][14]. Since governments, mass media, and many other organizations have high expectations for social media as information sharing tools, many organizations have opened official social media accounts[8, 10].

Twitter is both an information sharing tool and a form of social media, it is involved in many interactions among its users[5]. By analyzing interaction behaviors on Twitter, we can estimate how people use social media during crises. By clarifying how people use social media during disasters, we can encourage proper behavior on social media during the next disaster.

In this paper, we focus on the information sharing on Twitter during disasters. We collected data before and after the Great East Japan Earthquake that occurred at 14:46 on March 11, 2011. There are $362,435,649$ tweets, which were posted by $2,711,473$ users. Also, we used a dataset comprised of a network of followers that directly depicts a social graph of Twitter. The information of the dataset, which includes the follower/followee data of about 1 million Twitter users, was crawled in January 2011.

In this paper, we define March 7 to 10th as "before the disaster" and March 11 to 15th as "after the disaster".

2. RELATED WORKS

Social media, including Twitter, are effective tools for crisis communication. Some research has analyzed how people used Twitter during crises [7, 3]. Mendoza et al. investigated the behavior of Twitter users during the 2010 Chile earthquake and characterized Twitter in the hours and days following it [9]. Heverin et al. analyzed its role during violent crimes [6]. Miyabe et al. surveyed the user trends

WWW 2013 Companion, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2038-2/13/05.

of Twitter after the Great East Japan Earthquake [11] and analyzed the interaction among Twitter users.

Other researchers observed events that happened in the real world using social media [2, 13, 16, 1]. For example, Yardi et al. is investigating the modes of spreading news and reports that Twitter helps spread news related to local areas more than news related to broad areas [17]. Sakaki et al. estimated an earthquake's epicenter using Twitter [12]. Yang et al. [16] investigated Twitter user behavior by focusing on replies and hashtags. Bakshy et al. [1] argued that "Weak ties may play a more dominant role in the dissemination of information online than currently believed" on Facebook.

Our research, which observed the status of the real world using social media during a crisis, is different from other research because we focus on the changes of user behaviors between normal and emergency situations.

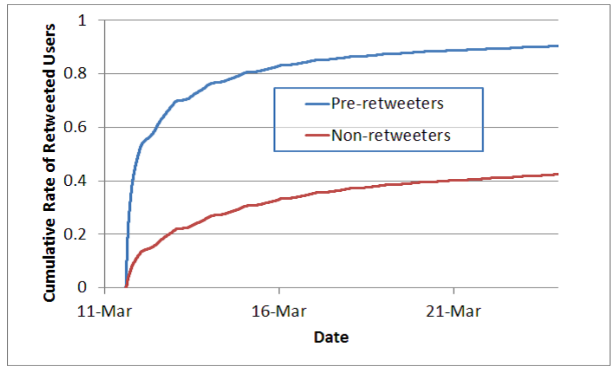

Figure 1: Cumulative rate of retweet users after disaster

3. REPLY AND RETWEET USAGE BEFORE AND DURING THE DISASTER

We analyze how well replies and retweets were utilized during the disaster. We classify the users into the following two groups. One is the pre-retweeters, who consisted of users who retweeted before the disaster. The other is non-retweeters; users who did not retweet before the disaster. The number of pre-retweeters was 520,302, and the number of non-retweeters was 2,191,171.

Figure 1 shows the cumulative rate of the users who retweeted after the disaster. The blue line shows the cumulative rate of the retweeted users of the pre-retweeters, and the red line shows the rate of the non-retweeters. Most of the users who retweeted before the disaster also aggressively retweeted after it. At the end of March 12, 69.0% of the pre-retweeters retweeted after the disaster. On the other hand, only 21.4% of the non-retweeters retweeted at the same time. At the end of March 23, which is the limit of our data, 52% of all users and 90.8% of pre-retweeters retweeted; however, only 43.1% of the non-retweeters retweeted. In other words, more than half of the users who had not retweeted before the disaster did not retweet after it.

We also compared users who used reply before the disaster (pre-repliers) and users who did not (non-repliers). At the end of March 12, 68.3% of pre-repliers and 25.5% of non-repliers used reply after the disaster. At the end of March 23, 82.3% of all users, 92.9% of pre-repliers, and 68.1% of non-retweeters used reply.

These facts show that the users who had employed specific features of social media before the disaster continued to actively use them during the disaster. On the other hand, not all of the users who had not previously used such specific features utilized them after the disaster.

4. REPLY AND RETWEET ON THE FOLLOWER NETWORK

To clarify why people used Twitter, we must analyze who interacted with whom. When people only interacted with their friends (followers and followees), they were using Twitter as a communication tool. On the other hand, interaction with non-friends suggests that Twitter was being used as an information collection tool. In this section, we classify each reply and retweet based on whether interaction occurred on the follower network.

Since following other users indicates a follower who is paying attention to a followee, on the follower network, interaction is with users who received much attention. In this paper, the follower data we used were collected before the disaster; the follower relationship was established based on the interest of normal days. In other words, out of follower network interaction possess properties that differ from the interaction of normal days.

Figure 2 shows the rate of the on the follower network replies and the retweets of all replies and retweets. Before the disaster, the percentage of on the follower network replies was roughly 32%, which only slightly changed during the disaster, showing that communication between friends did not change significantly during the disaster. In fact, 72.3% of the replies that occurred during the disaster were performed on the relation that had replied before the disaster. Communication between individuals like replies was conducted using a structure that existed before the disaster.

The percentage of the on the follower network retweets before the disaster was roughly 23%, which decreased to roughly 10% during the disaster. In other words, most of the information was shared over the follower network, because the information required under such a crisis situation was completely different from the information of normal days. Because there was no contact between the informants and the retweeters, the latter had no background knowledge about the former. In such a case, the retweets could only be decided from the global values of the information but not from private interests. Thus, out of follower network retweets are estimated to be more valuable information. The increase of such retweets should lead to the global diffusion of information beyond local relationships.

5. INFORMATION CLASSIFICATION

5.1 Retweet Clustering

In this section, we classify information that was diffused widely on Twitter to clarify what kind of information is required by victims. We used a bipartite graph [15] that consisted of tweets and retweeted users to classify the retweeted tweets. We estimated that the tweets that were retweeted by identical users share similarity, and thus pairs of tweets with such relationships are connected by a link to create a network of similar tweets.

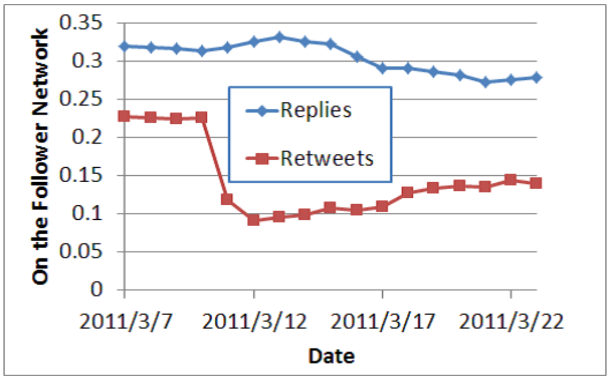

Figure 2: Rate of replies and retweets on follower networks

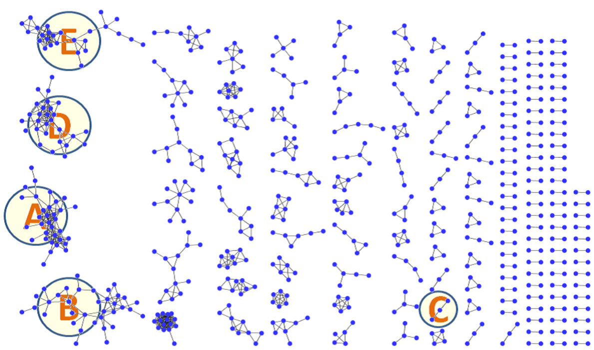

Figure 3: Retweet network

In this analysis, we used tweets that were retweeted more than 1000 times to analyze tweets that were diffused over a certain degree. First, we classified tweets by users who retweeted them to summarize similar tweets. For the pair of tweets rt_i, rt_j, we calculated how many users had retweeted both of these tweets. We represent a group of users who retweeted rt_i as U_i. If there are many overlapping users between U_i and U_j, we consider rt_i and rt_j an adjacent retweet. We constructed a network by extracting the adjacent retweets from our dataset.

We represent the overlapping degree of the retweeting users between U_i and U_j as O_{ij} and apply the Jaccard coefficient[4] to O_{ij}:

$$O_{ij} = \frac{|U_i| \cap |U_j|}{|U_i| \cup |U_j|} \quad (1)$$

If O_{ij} exceeds threshold value th, we link rt_i and rt_j and construct a network of tweets based on the similarity of the retweeting users. In this research, we chose $th = 0.04$, which maximized the number of network components with more than two nodes.

5.2 Types of Diffused Information

We found 168 components in the network. The network constructed by these processes is shown in Figure 3. We investigated the components with more retweets as examples to clarify what kind of information was diffused during the disaster from the viewpoint of retweet user similarity.

We confirmed the contents of the tweets that were included in the top five largest components shown as A-E in Figure 3. The contents of the components are shown in Table 1.

The tweets in component A were information calling for attention from mass media and public accounts. Component B also had mass information calling attention to the earthquakes that was tweeted by non-public users. Component C consisted of tweets that warned of harmful rumors about the Fukushima nuclear power plant accident. Component D consisted of information about planned blackouts that was tweeted by public accounts, such as mass media and local governments. Component E included information about the nuclear power plant disaster and radioactivity.

From the viewpoint of the tweet contents, these components can be classified as follows. The contents of components A and B are information calling for attention. The contents of components C and E are about the threat of nuclear accidents. The contents of component D contain information about electricity shortages. All of this serious information was urgently required during the disaster.

6. CONCLUSION

We analyzed 360 million tweets that were posted before and after the Great East Japan Earthquake to elucidate how people share the information on Twitter during disasters. We arrived at the following conclusions:

- Many users with little experience with such specific functions as reply and retweet did not continuously use them after the disaster.
- Retweets were well used to share information on Twitter.
- Retweets were used not only for sharing the information provided by general users but used for relaying the information from the mass media.

From our results, we conclude that social media users changed their behavior and reasons to autonomously use social media after serious events. People cooperated to change their Twitter mode from communication to an information sharing tool. This is the biggest advantage of Twitter during disasters. Of course, no certainties exist that Twitter will be used during subsequent disasters. Facebook may be the most useful social media during subsequent disasters, or maybe we will be using radically new social media. All social media, however, are expected to become useful tools for information sharing and communication, because of the changes in user behaviors during serious situations.

On the other hand, people rarely use a new function that they did not use before disasters. Therefore, we must encourage more users to use various functions of social media before serious events occur. However, excessive retweets sometime complicate searching for necessary information. Systems that easily organize and recommend important information during disasters are required.

In the future, we must observe and analyze the changes in social media to ascertain how such disasters affect social media over long periods. We must also clarify how best to exploit the information from social media in efforts to recover from disaster. Finally, another important item on our agenda is discussing how social media can be utilized to realize a resilient society.

Table 1: Spread information by retweets

	Nodes	Retweets	Contents
A	33	72893	Information calling for attention (media)
B	35	64407	Information calling for attention (general users)
C	3	54424	Caution about harmful rumors of radiation exposure
D	28	42129	Information about planned outages (from mass media)
E	26	38073	Information about radiation

7. ACKNOWLEDGMENTS

We thank Genta Kaneyama (Cookpad Inc.) for assistance in collecting data from Twitter. This work was supported by JSPS KAKENHI Grant Number 24300064 and Microsoft Research Asia University Relations.

8. REFERENCES

[1] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic. The role of social networks in information diffusion. In WWW2012, pages 519–528, 2012.

[2] 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.

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

[4] W. B. Frakes and R. Baeza-Yates. Information Retrieval: Data Structures and Algorithms. Prentice Hall PTR, 1992.

[5] M. Graves. The 2010 world cup: a global conversation. Twitter Blog, 2010. http://blog.twitter.com/2010/07/2010-world-cup-global-conversation.html.

[6] 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, 2010.

[7] 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.

[8] O. G. Initiative. Open govenment. http://www.whitehouse.gov/open.

[9] 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. ACM Press, July 2010.

[10] Ministry of Internal Affairs and Communications, Japan. WHITE PAPER Information and Communications in Japan. Japanese Government, 2012.

[11] 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.

[12] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In Proceedings of the 19th international conference on World wide web, WWW '10, pages 851–860. ACM, 2010.

[13] 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.

[14] S. Vieweg. Microblogged Contributions to the Emergency Arena: Discovery, Interpretation and Implications. In Computer Supported Collaborative Work, Feb. 2010.

[15] S. Wasserman and K. Faust. Social network analysis: Methods and applications. In Structural Analysis in the Social Sciences, volume 8, pages 299–302. Cambridge University Press, 1994.

[16] L. Yang, T. Sun, M. Zhang, and Q. Mei. We know what @you #tag:does the dual role affect hashtag adoption? In WWW2012, pages 261–270, 2012.

[17] 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.