Use Trend Analysis of Twitter after the Great East Japan Earthquake

Mai Miyabe
Center for Knowledge Structuring,
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo
113-8655, Japan
mai.miyabe@gmail.com

Asako Miura
Department of Psychological Science, Kwansei Gakuin University
1-1-155 Uegahara Nishinomiya,
Hyogo 662-8501 JAPAN
asarin@kwansei.ac.jp

Eiji Aramaki
Center for Knowledge Structuring,
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo
113-8655, Japan
eiji.aramaki@gmail.com

CSCW'12, February 11–15, 2012, Seattle, Washington, USA.
ACM 978-1-4503-1051-2/12/02.

Abstract
After the Great East Japan Earthquake in Japan 2011, numerous tweets were exchanged on Twitter. Several studies have already pointed out that micro-blogging systems have shown potential advantages in emergency situations, but it remains unclear how people use them. This paper presents a case study of how people used Twitter after the Great East Japan Earthquake. First, we gathered tweets immediately after the earthquake and analyzed various factors, including locations. The results revealed two findings: (1) people in the disaster area tend to directly communicate with each other (reply-based tweet). On the other hand, (2) people in the other area prefer spread the information from the disaster area by using Re-tweet.

Keywords
Micro blogging, disaster response, twitter, earthquake

ACM Classification Keywords
H.5.3 Group and Organization Interface – Collaborative computing, Computer-supported cooperative work, Web-based interaction; K.4.2 Social Issues

General Terms
Human Factors

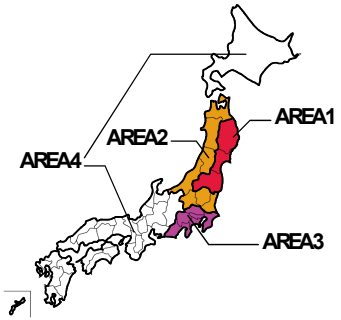

figure 1. Area Definition. We divide Japan into four areas. These boundaries come from a disaster level.

Introduction

Information and Communications Technologies (ICTs) play an important role in disaster situations. Among the many ICTs, micro-blogging systems, especially Twitter, have shown potential advantages because they are available with use of multiple devices, including cell phones; moreover, they are easy to use. In fact, people actively provided real-time situation updates using Twitter during various crises, such as the 2007 California Wild Fires [1, 2], the 2009 Red River Floods [3], and the 2009 Oklahoma Grassfires [4].

This study elucidates how people used Twitter immediately after a major disaster: the 2011 Great East Japan Earthquake. In fact, the number of re-tweets increased dramatically (up to 20 times higher than normal) immediately after the earthquake, reflecting the huge amounts of information that were being transmitted around Japan.

Earthquake corpus

From 11 March 2011 at 16:10 (1.5 hours after the earthquake), we started to gather Japanese tweets via Twitter API. Because the Twitter API does not support full tweet gathering, we specifically examined earthquake-related tweets, using a keyword, "earthquake". This keyword-based crawling yielded 1,612,074 tweets (by 30 March).

Area analysis

To detect the location of a tweet, we used a two-stage method: (1) GPS-based detection and (2) Address-based detection. We divide Japan's prefectures into four areas, which are shown in Figure 1.

- ***AREA1***: Disaster Area, including Fukushima, Miyagi, and Iwate Prefectures.
- ***AREA2***: Surrounding Area, including prefectures adjacent to Area 1.
- ***AREA3***: Eastern Japan Area, consisting of prefectures with power supplied by the Fukushima nuclear power plants.
- ***AREA4***: Other areas.

Analysis on Re-tweet

Response protocols of two types are used in Twitter: (1) an official re-tweet and (2) an unofficial re-tweet. Although the official re-tweet has been supported since 2010, the unofficial one was used long before. The usage of re-tweets of both types shows that official re-tweets are far fewer than unofficial re-tweets. Therefore, this paper only addresses unofficial re-tweets. To detect an unofficial re-tweet, we extracted tweets that contain "RT."

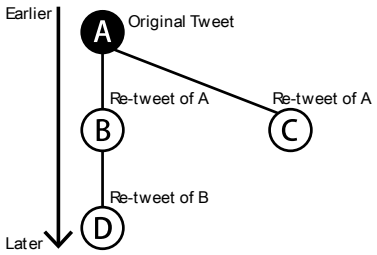

figure 2. Original tweet and its Re-tweets. Tweet (A) has two re-tweets: (B) and (C). Tweet (B) has one re-tweet: (D). From this situation, three pairs are extracted: (Original: Re-tweet) = (A: B), (A: C), and (A: D). We always deal with a relation between the re-tweet and the original tweet.

In order to examine the information flow, we investigate area information of two types: (1) the original tweet area and (2) a re-tweet area. To detect the original message and its area, we retrieved the same message to each re-tweet message. We regard the earliest time-stamped message as the original message (Figure 2).

Results

Finding 1: people in the disaster area prefer reply-tweet than the other areas.

Table 1 shows the number of tweets, re-tweets, and replies. Figure 3 shows a map of re-tweet ratio and reply ratio. These results indicates that re-tweet ratio of the disaster area (AREA1 and AREA2) was lower than that of the other area. Instead, such area relies mainly on direct communication (reply).

table 1. Number of tweets, re-tweets, and replies.

	Number of tweets (A)	Number of RT (B)	# of Average RT (B/A)	Number of reply (C)	# of Average reply (C/A)
AREA1	51,791	12,963	25.0%	10,371	20.0%
AREA2	60,097	17,327	28.8%	11,238	18.7%
AREA3	497,831	166,355	33.4%	85,480	17.2%
AREA4	221,961	93,176	42.0%	34,982	16.5%

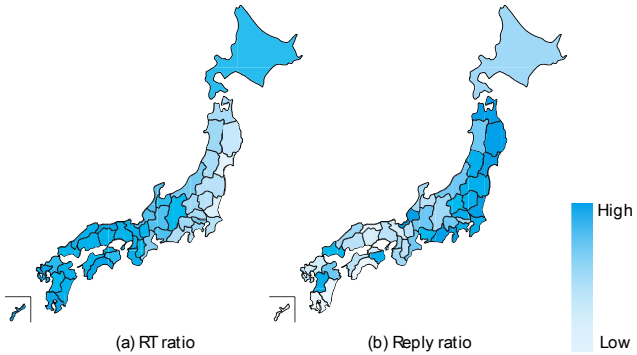

figure 3. re-tweet ratio (Left) and reply ratio (Right).

Finding 2: information posted from the disaster area tends to spread in the other area.

We specifically examine the information flow in the emergency situation. We define the tweet-transfer rate (TTR) as follows:

TTR= (number of re-tweets that originate in the area, and which are re-tweeted in the other area) / (number of re-tweets that originate in the area)

table 2. Re-tweet flow (Original tweet and Re-tweet area).

		Re-tweet area			
		AREA1	AREA2	AREA3	AREA4
Original tweet area	AREA1	12,470	3,136	26,323	14,989
	AREA2	1,430	10,517	19,527	13,160
	AREA3	26,239	45,042	527,516	280,054
	AREA4	6,047	10,778	99,789	110.062

Table 2 shows the number of tweets classified according to original areas and re-tweet areas. From this table, most of tweets transmitted from AREA 1 were re-tweeted in the other areas. Similarly, those from AREA 2 were re-tweeted in the other areas.

Figure 4 shows a time-line chart of TTR. From this figure, TTRs of AREA1 and AREA2 were higher than those of AREA3 and AREA4. Most of the tweets posted from AREA1 and AREA2 were re-tweeted in the other areas. Because AREA 1 and AREA 2 are the worst-affected areas by the earthquake, we can say that information from a disaster area tends to be transferred to the other areas.

Conclusion

This study conducted a case study to investigate how Twitter was used immediately after the Great East Japan Earthquake. The results are two-fold: (1) people in the disaster area tend to directly communicate with each other (reply-based tweet). On the other hand, (2) people in the other area prefer spread the information from the disaster area by using Re-tweet.

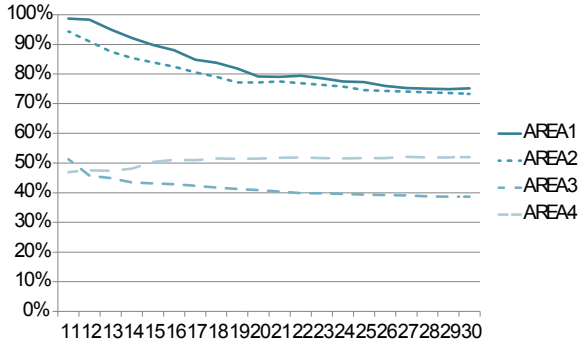

figure 4. Time-line chart of tweet-transfer rate.

References

[1] M. Glaser. California Wildfire Coverage by Local Media, Blogs, Twitter, Maps and More. PBS MediaShift, 2008.

[2] J. Sutton, L. Palen and I. Shklovski. Backchannels on the Front Lines: Emergent Use of Social Media in the 2007 Southern California Fires. In Proc. Information Systems for Crisis Response and Management (ISCRAM), 2008.

[3] K. Starbird, L. Palen, AL. Hughes and S. Vieweg. Chatter on The Red: What Hazards Threat Reveals about the Social Life of Microblogged Information. In Proc. Computer-Supported Cooperative Work (CSCW), 2010.

[4] S. Vieweg, AL. Hughes, K. Starbird and L. Palen. Microblogging during two natural hazards events: what Twitter may contribute to situational awareness. In Proc. Conf. on Human Factors in Computing Systems (CHI), pp. 1079-1088, 2010.