Can Twitter catch precursory phenomena before sediment disasters?

Masaru KUNITOMO, master[1]; **Joko KAMIYAMA, master**[1]; Kazuki MATSUSHITA, master[1]; Yuzuru YAMAKAGE, doctor[2]; Kunihiro TAKEDA, master[2]; Cheng QIU, doctor[3]; Akiko ITO, bachelor[3]; Takeru ARAKI, master[3]

ABSTRACT

In Japan, the Cabinet Office recommends that municipalities immediately judge issuance of evacuation instructions when they grasp the signs of sediment-related disasters. However, this early warning mechanism has not necessarily functioned effectively, because it is difficult for local governments to get such information from local residents. In such situation, "twitter" has been sharply growing in the number of users and is being incorporated into social system. Then, with focus on the real-time property of "twitter", the authors have started a research to identify information on the precursory phenomena of sediment-related disasters contained in users' "tweets". Then, we confirmed that tweets are fairly effective in detecting precursory phenomena etc. particularly in areas with a large population size. And that application to early warning systems can be expected.

KEYWORDS

Sediment Disaster; early warning; SNS; Twitter

INTRODUCTION

It has been said that signs of sediment-related disasters, such as "earth rumbling" and "earth smelling," are important early warning signal, and the Cabinet Office of Japan recommends on its guideline that municipalities immediately judge issuance of evacuation instructions and local residents quickly take actions to protect their safety when they grasp the signs. In fact, there are many cases reported where local residents found precursory phenomena of sediment-related disasters and evacuated together with family members and/or neighboring residents to avoid personal injuries. However, there are also reports that even if local residents find precursory phenomena of sediment-related disasters, such information is transferred only to family members and/or neighboring residents, and is rarely reported to local governments (Miyase et al., 2009). Thus, information on the precursory phenomena of sediment-related disasters is effective in judging evacuation, but there are some issues to be solved concerning quick collection and sharing of such information to use such phenomena information for early warning system.

1 National Institute for Land and Infrastructure Management MLIT, Tsukuba, JAPAN, kamiyama-j253@nilim.go.jp

2 Fujitsu Laboratories Ltd.

3 Nippon Koei Co. Ltd.

In such situation, "twitter" has been sharply growing in the number of users with the advantages of real time and spreading properties and is being incorporated into social system. In disaster prevention as well, cases of using twitter as a risk communication tool has been increasing.

Then, with focus on the real-time property of "twitter" information, the authors have started a research to identify information on the precursory phenomena or occurrences of sediment-related disasters contained in users' "tweets" stating uncertainties or fears about heavy rain etc. and incorporate them into early warning systems.
In this study, we analyzed the tweets posted in the events of the July 2012 Northern Kyushu Heavy Rain and the August 2014 Hiroshima Heavy Rain and studied the possibility of detecting the situation of the site where a sediment-related disaster occurred using twitter information.

CHARACTERISTICS OF TWITTER

Twitter is a kind of social media called "mini blog" or "micro blog." As characteristics of twitter, Kazama (2012) states that twitter information quickly and widely spreads due to the high real-time property despite the limitation to a maximum of 140 characters per tweet, easier information exchange with or transfer to other users, and more casual and easier communication of information among twitter users than general SNS media.
There are many cases reported where twitter was very effective due to its real time and spreading properties in risk communication immediately after the occurrence of disasters (e.g., Ishikawa et al, 2012). On the other hand, it is also fact that some are concerned about spread of false rumors.

Thus, there seems to be pros and cons on using twitter as a risk communication tool. However, Taniguchi (2012) points out that twitter is effective as an information sharing tool in the initial phase of a disaster where it is difficult to grasp the overview even if negative effects such as spread of false rumors are disseminated.

Furthermore, in recent years, studies to probe the possibility of using twitter as a social sensor have been also going on, including estimation of the epicenter of an earthquake by analyzing tweets and detection of the spread of influenza (e.g., Sakaki et al., 2010).
As we reviewed the findings of such prior studies comprehensively, it is considered that use of twitter as a social sensor enables detection of natural phenomena that are hardly detected by physical sensor, although the reliability and stability of twitter are much lower than those of physical sensors, and that twitter can be a useful tool to visualize what cannot be detected by conventional means by supplementing physical sensor information.

EXAMINATION OF THE POSSIBILITY OF GRASPING THE CONDITION OF THE DISASTER SITE

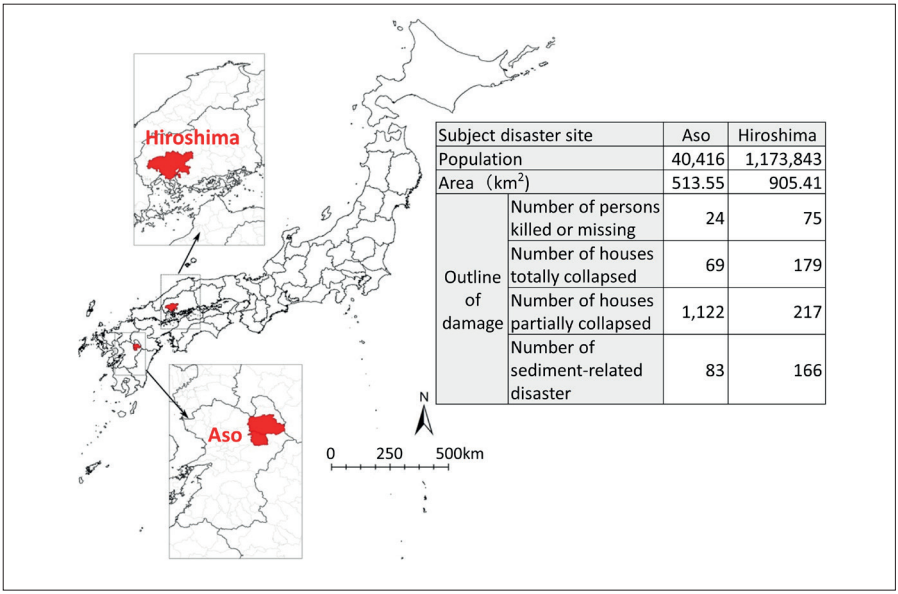
ANALYTICAL METHOD

After screening the twitter information on the Web to identify the municipalities where the occurrence of a disaster is referred to in posted tweets ("municipality estimation"), we extracted tweets related to heavy rain and classified them into the situation categories such as sediment-related disasters and weather condition etc. in order to grasp the situation of the site with tweet content of such extracted information and set key words that represent each situation category.

Then, we extracted tweets that include a key word we set up and estimated the situation based on changes in the number of tweets in each situation category. Further, we examined the effectiveness of grasping the situation from the tweets posted with critical content suggesting an emergent situation. To carry out municipality estimation, we adopted the procedure taken by Takeda et al (2014) through partial simplification.

SUBJECT DISASTER AND OUTLINE OF TWITTER INFORMATION

This study examined the sediment-related disasters caused by the Northern Kyushu Heavy Rain and the Hiroshima Heavy Rain, which hit Minami-Aso village, Aso city, Kumamoto prefecture ("Aso") and Hiroshima city, Hiroshima prefecture ("Hiroshima"), respectively, and in both of which it is known that local residents found precursory phenomena of sediment-related disasters (e.g., Sakai et al., 2013) (Figure 1). Time zones selected to analyze data were



Subject disaster site		Aso	Hiroshima
Population		40,416	1,173,843
Area (km^2)		513.55	905.41
Outline of damage	Number of persons killed or missing	24	75
	Number of houses totally collapsed	69	179
	Number of houses partially collapsed	1,122	217
	Number of sediment-related disaster	83	166

Figure 1: Locations and outline of subject disaster sites

from 16:00 on July 11 to 12:00 on July 12, 2012 for the Northern Kyushu Heavy Rain and from 1:00 to 5:30 on Aug. 20, 2014 for the Hiroshima Heavy Rain. In both events, rainfall rapidly increased at midnight and sediment-related disasters occurred at dawn and caused damage etc. to many people.

Of the tweets posted during the duration for analysis above, 2,207 tweets were estimated to have been posted from Aso, and 5,813 tweets, from Hiroshima.

EXTRACTION AND CLASSIFICATION OF TWEETS RELATED TO DISASTERS

Of the tweets of which locations were estimated, we extracted 968 tweets related to the Aso and 1,617 tweets for the Hiroshima, based on the contents of tweets. Further, the extracted tweets were classified into the "situation categories" of Table 1.

Table 1: Situation category.

Situation categories	Contents
Weather condition	Weather condition such as downpours, thunder, etc.
Warning/Evacuation/ Rescue activities	Announcement of the weather warning, official announcement such as evacuation instruction, and rescue activities.
River condition	River condition such as swelling torrent, muddy stream.
Flood disaster	Flood disaster such as overflow and inundation.
Sediment –related disaster	Sediment-related disaster such as debris flow and landslide.
Closure of school/Closure	Things related to closure of school and closure.
Traffic condition	Relation to the traffic condition such as road blocked and delay of transportation.
Feeling/State of mind	Anxiety to disaster occurrence.
Others	Not categorized as above.

KEY WORD SETTING BY TEXT ANALYSIS ACCORDING TO SITUATION CATEGORIES

Using "KH Coder," software for text analysis, we automatically extracted frequent words according to situation categories by analyzing the tweets after decomposing them into part of

Table 2: Examples of key words (Category related to sediment-related disasters).

Aso			Hiroshima		
Key word	Part of speech	Number of appearance	Key word	Part of speech	Number of appearance
Landslide	Noun	34	Landslide	Noun	107
Buried alive	Noun	20	Sediment	Noun	29
Sediment	Noun	8	Disaster	Noun	26
Disaster	Noun	5	Collapse	Verb	11
Debris flow	Noun	3	Hill behind	Noun	9

speech. From among the words extracted, we adopted as key words representing the content of each situation category (e.g. "land slide", "disaster", etc. for "Sediment-related disasters" and "Dangerous", "Scary", etc. for "Feeling and state of mind."). Table 2 provides examples for the category of "Sediment-related disasters."

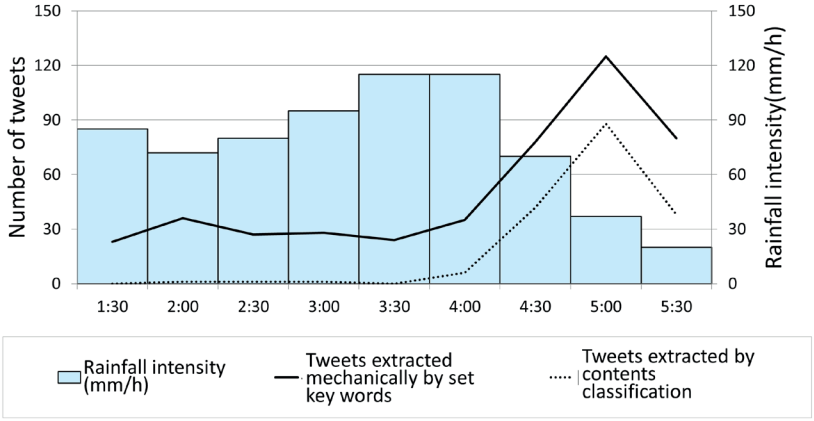


Figure 2: Time series changes in the number of tweets in the category of sediment-related disasters (Hiroshima)

Figure 2 shows in graph form time series changes in the number of tweets of which contents were classified and extracted after text analysis for the category of sediment-related disasters in the Hiroshima Heavy Rain (, which do not necessarily contain key word(s)) and the numbers of tweets extracted mechanically using the set key words, together with changes in intensity of rainfall. Since the trends of both types of numbers were generally consistent, it was verified that the key words set up can be used to capture changes in the number of tweets in each situation category.

GRASP OF THE SITUATION USING KEY WORDS

Figure 3 shows the timing of response actions etc. to the disaster in Aso and rainfall, changes in the number of tweets extracted using key words for each situation category, and the timing of appearance of tweets showing critical content.

The number of tweets in the weather related category is increasing continuously, which suggests the situation where heavy rainfall is continuing. Meanwhile, tweets in the categories of flood and sediment-related disasters begin to increase about 7:00, which is far behind the time of the first reporting of the disaster. However, when we checked the contents of tweets individually, we found that the occurrence of a sediment-related disaster could be detected a little earlier (6:22). Additionally, when tweets are not limited to those for which the location

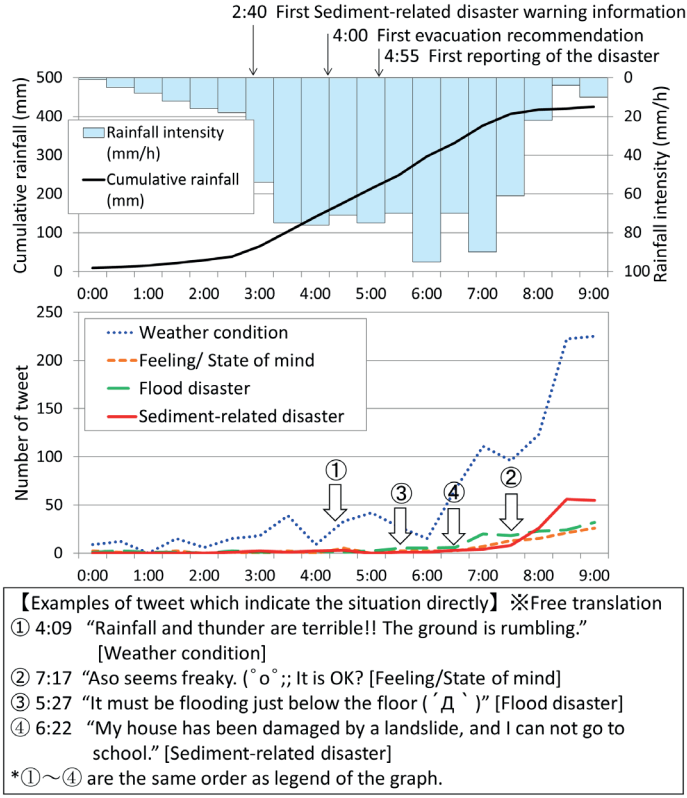


Figure 3: Grasped disaster situation by analyzing tweets (Aso)

is estimated, relevant tweets appeared in a range of heavy rain areas including Aso in much earlier time zones, e.g., "Landslide, now" at the time between 2:00 and 3:00.

Figure 4 provides the results of analysis conducted on the Hiroshima Heavy Rain. For Hiroshima, changes in the number of tweets and contents of tweets also suggest the situation where heavy rainfall is continuing. This is similar to Aso, but the number of tweets in the category of sediment-related disasters increased in time zones much earlier than the occurrence of the sediment-related disaster, which is different from the situation in Aso. Further, in Hiroshima, when tweets are not limited to those for which the location is estimated, tweets stating the detection of precursory phenomena appeared at 2:50, earlier than the time zone during which debris flow is said to have occurred intensively, such as "I've been hearing a sound like the one heard when stones are rolling down the river" (this was estimated since heavy rain was falling only in Hiroshima throughout the Japan when the tweet was posted). Difference of the size of population (the number of twitter users) is

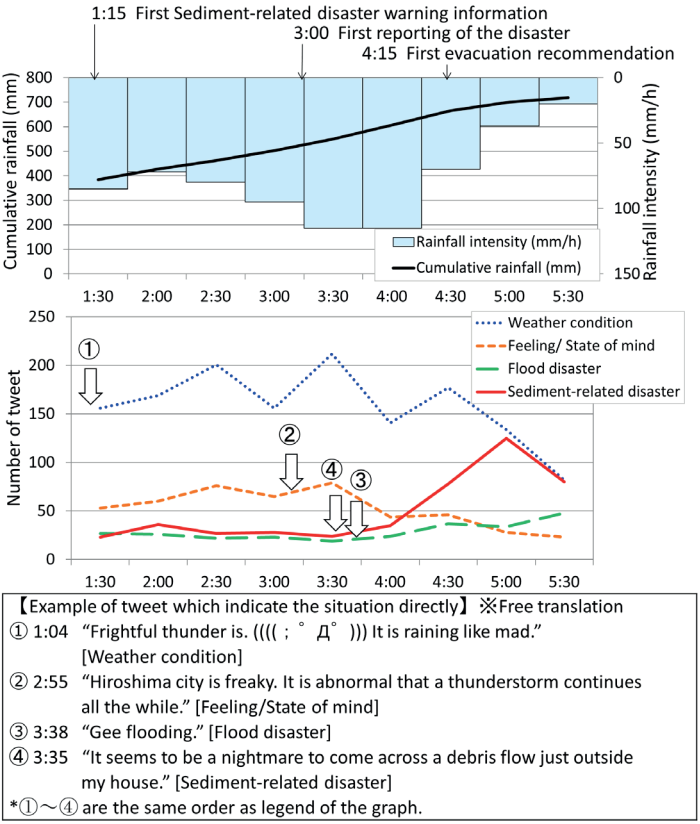


Figure 4: Grasped disaster situation by analyzing tweets (Hiroshima)

clearly the main factor of the early increase in the number of tweets in Hiroshima compared to Aso. Therefore, in order to detect precursory phenomena etc. at an earlier phase, it is important to extract tweets including critical content individually, as well as changes in the number of tweets. At present, when using twitter as an early warning system, it is necessary to provide measures to prevent omission of critical tweets, such as joint use of tweets of which locations are limited to prefectural level.

CONCLUSIONS

This study examined the possibility of grasping precursory phenomena from changes in the number of tweets extracted using key words and from the timing of appearance of tweets including critical content by setting key words after classifying tweets related to the heavy rain into situation categories. As a result, we confirmed that tweets are fairly effective in detecting precursory phenomena etc. in areas with a large population size, such as Hiroshima. It was also found that information on the feelings, state of mind, etc. of local residents, which

can be of any help in determining directions for evacuation, can be fairly grasped and that application to warning / evacuation systems can be expected.
On the other hand, in areas with a small population, it is difficult to detect precursory phenomena etc. only with the increase in the appearance of tweets since the number of tweets is very small. Therefore, in order to grasp the precursory phenomena of sediment-related disasters, etc. using twitter information, including areas with a small population, we intend to grasp situation changes sensitively by introducing the technology for evaluating the value of each tweets according to areas and to reduce omissions of tweets that include critical content by improving the accuracy of location estimation.
We also intend to develop application software to improve the reliability as disaster prevention information, e.g. by overlapping display of the heavy rainfall areas obtained from the precipitation radar.

REFERENCES

- Masayuki M., Toshihiro K., Shota K. (2010) Residents' Attitude Survey in the July 2009 Yamaguchi Heavy Rain Disaster, FY 2010 Japan Society of Erosion Control Engineering Presentation Summary
- Kazuhiro K. (2012) Information Dissemination in Twitter, Journal of the Japanese Society for Artificial Intelligence, Vol. 27, No. 1: 35-42.
- Ikki O., Yutaka M. (2012) Twitter and social media, Journal of the Japanese Society for Artificial Intelligence, Vol. 27, No. 1: 34.
- Tetsuya I., Akiyuki K., Kimiro M. (2012) Investigation of information sharing by Twitter users during the 2011 heavy snow disaster in San-in region of western Japan, A Collection of Papers of the Institute of Social Safety Science No. 17. http://isss.jp.net/isss-site/wp-content/uploads/2013/08/2012-014_cd.pdf
- Shinichiro T. (2012) Effectiveness of twitter at disasters --- Taking as an example the heavy rain disaster by Typhoon No. 12 in September 2011, Disaster Information No. 10: 56-67.
Sakaki, T., Okazaki, M., Matsuo, Y. (2010) Earthquake shakes twitter users: Real-time event detection by social sensors, Proc. 19th Conf. on World Wide Web, 851-860.
- Kunitaka T., Tadanobu F., Shigetaka T., Yuzuru Y., Ken A., Akiko I., Ken M., Junichi K., (2014) Estimation of Occurrence of Disaster Using Twitter Data --- Report on the cases investigated concerning sediment-related disasters in the 2012 Northern Kyushu Heavy Rain, FY2014 Japan Society of Erosion Control Engineering Research Presentation Summary B: 172-173.
- Nobuaki S., Ryoichi M., and Toshihiro K. (2013) Behaviors of residents in case of sediment-related disaster caused by the July 2012 Northern Kyushu Heavy Rain, Journal of the Japan Society of Erosion Control Engineering (Shin-Sabo), vol. 66, No. 2: 57-63.