

Disaster Warning Map using Premonitory Symptom Keywords Gathered from Twitter in Real Time

Otsuki A^{1*}, Kuwana A² and Kawamura M³

¹Nihon University, 1-3-2 Misaki-Cho, Chiyoda-Ku, Tokyo 101-8360, Japan

²Ochanomizu University, 2-1-1 Otsuka, Bunkyo-Ku, Tokyo 112-8610, Japan

³MK Future Software, Ibaraki, Japan

Abstract

In this research project, the premonitory symptom keywords and location information relating to landslides, tsunamis, river flooding and sudden downpours of rain were gathered from Twitter (SNS Data) in real time. The data was analyzed, a hazard coefficient was calculated, and this coefficient was then visualized on a heat map so that a disaster warning map could be developed. In 2016, at the time of the occurrence of Typhoon No. 16 in Japan, about 4000 Tweets were acquired and an experiment was carried out for the evaluation of river flooding. Specifically, based on the results verified by the summarized damage reports from Typhoon 16 announced by the Japanese Cabinet Office (report) and on the relevant news reports, it was clearly possible to predict these events more quickly by using the disaster warning map than by relying on the news reporting services for about 57% of the actual occurrences (21 instances) where the water surfaces of rivers went over the danger level.

Keywords: Disaster information studies; SNS data analysis; Words for premonitory symptoms; Hazard map

Introduction

Japan is an island country and as a result of its climatic conditions a variety of disasters can occur; therefore, disaster prevention is very important and a variety of disaster prevention policies have been implemented by institutions including the Japanese Cabinet Office [1,2]. Recently, the Ministry of Land, Infrastructure, Transport and Tourism, (“MLIT”) has used Tweeted information to gather the premonitory symptoms of landslides [3], thus showing a move toward using SNS in disaster prevention measures. In their research, the words relating to the premonitory symptoms of landslides, tsunamis, river overflows and sudden downpours were used as keys and Tweets, while the location information was acquired in real time. Then, in the areas (municipalities) where the number of premonitory symptom words passed a predetermined threshold, hazard coefficients were calculated and a disaster warning map was developed where these hazard coefficients were displayed in the form of a heat map. In the MLIT example, only landslides were recorded, but the current research has been expanded to include tsunamis, river overflows, and downpours of rain. In addition, the warning information (hazard coefficient) is visualized on a heat map, which is a further difference from the MLIT project. In the evaluation experiment, about 4000 Tweets that occurred during Typhoon No. 16 in 2016 were gathered. Once the Cabinet Office reports [4] were checked for overflowing rivers and the results were verified based on the relevant news reports, it was discovered that the warning map could clearly predict 57% of the actual occurrences (21 instances) of river overflows or cases of the water surfaces of rivers rising over the danger level before the time of the news reports.

Proposed Method

Process leading to creating the disaster warning map

In this research, a method of creating a disaster warning map to warn of a danger before the disaster develops is proposed. The process leading to this is as follows.

① Gather Tweets in real time that include the premonitory symptom keywords and aggregate them by municipality units.

↓

② Calculate a HAS (Hazard Alarm Score, see next Section per municipalities where the number of Tweets that include the premonitory symptom keywords exceeds 0.01% of the population of the municipality.

↓

③ Visualize the HAS as a HEAT map.

To gather the real time Tweets, a Twitter Streaming API was used. The Tweeted municipality was included as the full name item [5] of the places that can be acquired by the Twitter Streaming API, and based on these items, the Tweets were aggregated by the municipality.

The trigger to calculate the HAS (the HAS threshold) was determined as follows: the average population of Japanese municipalities is about 78,000, according to the population preliminary aggregate result in the 2015 census [6]. It can be expected that there will be a few cases where the number of Tweets that include the premonitory symptom keywords will go over 5 in a municipality with no danger of disaster, so the number of Tweets is given through the formula [$5/78,000 \approx 0.01\%$], where the value of 0.01% or more of the municipality population acts as a trigger to calculate the HAS. In addition, as is shown by “t” on the horizontal axis in Figure 1, the trigger check operates over a space of 10 minutes. This image shows Municipality “A” from the time of the first check (that is t) and Municipality “B” from the third check (t+20), with

*Corresponding author: Otsuki A, Associate Professor, College of Economics, Nihon University, Chiyoda-Ku, Tokyo, Japan, Tel: 03-3219-3480; E-mail: otsuki.akira@nihon-u.ac.jp

Received January 27, 2018; Accepted February 15, 2018; Published February 23, 2018

Citation: Otsuki A, Kuwana A, Kawamura M (2018) Disaster Warning Map Using Premonitory Symptom Keywords Gathered from Twitter in Real Time. J Geogr Nat Disast 8: 217. doi: [10.4172/2167-0587.1000217](https://doi.org/10.4172/2167-0587.1000217)

Copyright: © 2018 Otsuki A, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

each HAS calculated. Figure 1’s diamond shapes and square shapes each represent Tweets that include the premonitory symptom keywords, and reveals that at the 4th check time (t+30), Municipality A has a total of 18 individual Tweets accumulated to “t+30”. The HAS is then renewed (recalculated) using these accumulated Tweets.

Next, the premonitory symptom keywords used in the present evaluation experiment are shown in Table 1. As is shown in Section the Results of the Evaluation Experiment of this research, the evaluation experiment was carried out for the occurrence of overflowing rivers, so Table 1 shows the premonitory symptom keywords relating to this type of disaster.

Calculating the HAS (Hazard Alarm Score)

The formula for calculating the HAS (Hazard Alarm Score) is shown in the following equation. The bigger the HAS score displayed, the higher the danger of a disaster will be occurring.

$$HAS_i = \frac{NT_i}{P_i} \times 100000 \quad (1)$$

Here, the “i” in “HAS_i” stands for the municipality, NT_i (Number of Tweets) shows the number of Tweets with the premonitory symptom words, and P_i (Population) shows the population of the municipality. Assessing the hazard with only NT_i may produce a feeling of inequality due to the different populations, so in Equation 1, NT_i is divided by the population of the municipality. Normally, it is not the population, but rather the ratio of people covered by Twitter that is significant.

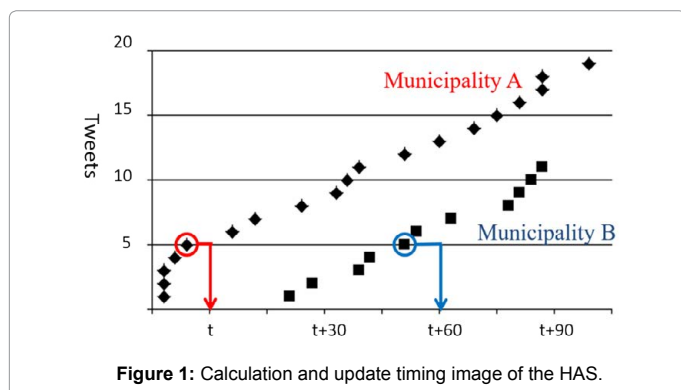


Figure 1: Calculation and update timing image of the HAS.

Targeted disaster [source]	Premonitory symptom keywords
Overflowing river [7]	flooding, inundation, rising water, sandbag

Table 1: Keywords for the premonitory symptoms of river floods.

There is no resource that reports the ratio of Twitter users in each of the Japanese municipalities; however, it has been reported that the number of Twitter users is roughly proportional to the population [8], so for the purpose of this experiment we divided the Tweets by the number of the municipal population. Also, the 100,000 at the end of the equation (1) is a value used for adjusting the HAS value to a figure that is easy to read as a natural number+a digit in the first decimal place, without any meaning of its own. For this evaluation experiment, it was set to 100,000 but as the number of targeted Tweets (NT_i) increases, this value will become smaller.

Evaluation Experiment

Outline of the evaluation experiment

The evaluation experiment was carried out with the premonitory symptom keywords listed in Table 1, as well as with about 4,000 Tweets gathered at the time when Typhoon No. 16 occurred in 2016. Figure 2 shows the aggregated Tweet numbers including the premonitory symptom words for each of the disaster items in Table 1. As these Tweets were acquired during the typhoon, and as virtually no Tweets were collected for sudden downpours, tsunamis or landslides, the effectiveness of the prediction map was tested only for river overflows.

Results of the evaluation experiment

Based on the Cabinet Office Report describing the damage resulting from Typhoon No. 16 of 2016 [4] and the news reports, we were able to clarify the time when the river surface exceeded dangerous levels. The number of Tweets before that time was used to calculate the HAS. On this occasion, there were virtually no cases of actual river flooding so the HAS calculation targeted the areas where the river surface passed the danger level for flooding. That is, on this occasion the HAS was calculated as “(a) the number of Tweets containing premonitory symptom words (before the news reports were broadcast)/(b) the population of the municipality*100,000”. The reason why (a) focused on the Tweets acquired before the news broadcasts is that the purpose of this evaluation experiment was only to verify the effectiveness. When the warning map is actually used, the focus will not be on the Tweets from before the news broadcasts, but the HAS will be calculated per municipalities where the number of Tweets including premonitory symptom words exceeds 0.01% of the population for that municipality.

In this research, a processing script was prepared to gather the Tweets and to extract the premonitory symptom keywords from the Tweets that were collected. Table 2 is an excerpt of the script output from this evaluation experiment. In Table 2, the “Tweet day and time” shows the day and time of the Tweet, where the time is converted to Japanese

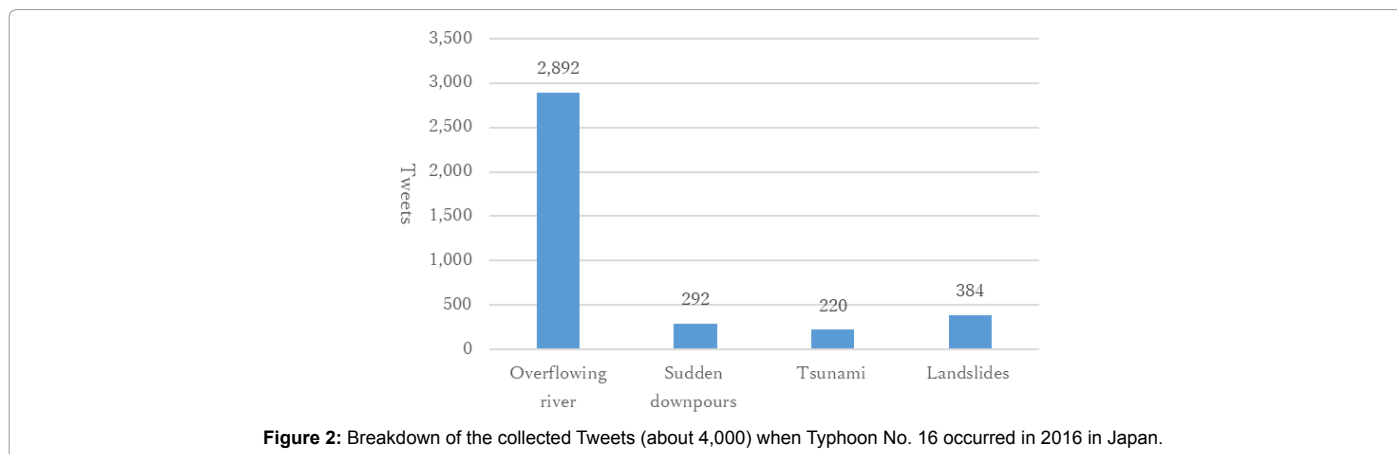


Figure 2: Breakdown of the collected Tweets (about 4,000) when Typhoon No. 16 occurred in 2016 in Japan.

Tweet day and time	Tweet texts gathered in real time	Premonitory symptom keywords about sudden floods				Municipality	(a) Number of Tweets	(b) Population	HAS ((a)/(b)*100,000)
		flooding	inundation	rising water	sandbag				
Tue Sep 20 10:46:28 +0000 2016	There was torrential rain from Typhoon No. 16 and for a while I didn't know what would happen, but the wind and rain have eased, and I feel better. In parts of the town there's an inundation above and below floor level and on the eastern side there is some damage from landslides in the hills. That's enough of typhoons!		1			Okazaki-shi, Aichi	20	381,031	5.248917
Tue Sep 20 09:23:22 +0000 2016	Just now, it looks like the eye of the typhoon. The wind and rain have eased, but there is rising water in the usually shin-level-deep river # Shinpukuji River @ Shiba Kyuen for grape picking.			1		Okazaki-shi, Aichi	19		
Abbreviated									
Mon Sep 19 14:53:21 +0000 2016	News about a shelter in Okazaki-shi! There are no landslides around here but there is flooding so it seems it has been opened. #Okasaki-shi.	1				Okazaki-shi, Aichi	1		

Table 2: Calculation process of the HAS and results of the HAS scores.

time so it can be compared with the times of the news broadcasts. The "Tweet texts gathered in real time" column includes the texts of the Tweets, and from this each of the premonitory symptom keywords is extracted in the Premonitory symptom keywords about "sudden floods" column, with the numbers put into the "flooding", "inundation", "rising water", and "sandbag" columns for each output. The municipality column shows the municipalities where the Tweets came from, while the "(a) Number of Tweets" column shows the NT_i of Equation 1. As an example from Table 2, in Okazaki-shi in Aichi Prefecture, there were a total of 20 Tweets that included the premonitory symptom words, so the value indicated in the "(a) Number of Tweets" column is 20. Note that even when there are multiple premonitory symptom keywords in 1 Tweet, the Tweet gets a 1 count. The "(b) Population" column shows the population of the municipality where the Tweet occurred, and the HAS column shows the HAS score calculation result: "(a) Number of Tweets column/(b) Population column*100,000".

Table 3 presents a summary of the above calculation results. Column 1 of Table 3 shows each municipality and its population, Columns 2 and 3 include the totals of the collected Tweets and the number of these Tweets acquired before the news broadcast, and Column 4 shows the HAS. In Table 3, the areas from Shimanto-shi in Kochi Prefecture to Takamatsu-shi in Kagawa Prefecture are the areas where hazard warnings were made using the HAS before there were news broadcasts. For the areas from Oita-shi in Oita Prefecture to Tsushima-shi in Nagasaki Prefecture, hazard warnings were not made before the news was broadcast.

Consideration of the Results of the Evaluation Experiment

Recall, precision and f-number

In this section, the Recall, Precision, and F-measure, which are shown in Equations 2-4, are calculated and the effectiveness of the

warning map is checked.

$$\bullet \text{ Recall} = \frac{NMH}{NMA} \quad (2)$$

• NMH (Number of Municipalities where the HAS was calculated)... Number of municipalities where the HAS was calculated

• NMA (Number of Municipalities where news Alerts were issued)... Number of municipalities where news alerts of river floods or dangerous water surface conditions were issued

$$\bullet \text{ Precision} = \frac{NMAH}{NMHR} \quad (3)$$

• $NMAH$ (Number of Municipalities where news Alerts were issued out of the Number of Municipalities where the HAS was calculated)... Number of municipalities where news alerts of river floods or dangerous water surface conditions were issued out of the number of municipalities where the HAS was calculated

• $NMHR$ (Number of Municipalities where the HAS was calculated Regardless of the news alerts)...Number of municipalities where the HAS was calculated Regardless of the news alerts of river floods or dangerous water surface conditions

$$F = \frac{(2 \times \text{Recall} \times \text{Precision})}{(\text{Recall} + \text{Precision})} \quad (4)$$

Using the data from Table 3, the Recall was calculated as "12/21=57.1%", the Precision was "12/12=100%", and the F-measure was "2*(0.571*1)/(0.571+1)=72.7%".

Summary of the discussion and future issues

This evaluation experiment was carried out based on materials sourced from the Cabinet Office Report about the Damage Resulting from Typhoon No. 16 of 2016 [4]. Since there were no municipalities

Name of the municipality/(b) population[6]	Number of Tweets including the premonitory symptom words		HAS
	Total	(a) Before the news broadcast	
Shimanto City, Kochi Prefecture/34,315	5	5	14.6
Takeo City, Saga Prefecture/49,108	5	5	10.2
Saiki City, Oita Prefecture/72,203	5	5	6.9
Okazaki City, Aichi Prefecture/381,031	26	20	5.2
Yokkaichi City, Mie Prefecture/311,089	16	16	5.1
Kuwana City, Mie Prefecture/140,226	5	5	3.6
Suzuka City, Mie Prefecture/196,251	6	6	3.1
Ichinomiya City, Aichi Prefecture/379,954	11	11	2.9
Toyota City, Aichi Prefecture/422,780	12	12	2.8
Kasugai City, Aichi Prefecture/306,599	8	8	2.6
Tokushima City, Tokushima Prefecture/258,602	7	7	1.9
Takamatsu City, Kagawa Prefecture/420,943	16	5	1.2
Oita City, Oita Prefecture	5	4	-
Hyuga City, Miyazaki Prefecture	10	2	-
Nobeoka City, Miyazaki Prefecture	3	3	-
Ogaki City, Gifu Prefecture	2	2	-
Nagoya City, Aichi Prefecture	2	2	-
Kagoshima City, Kagoshima Prefecture	2	1	-
Anan City, Tokushima Prefecture	1	1	-
Miyakonojo City, Miyazaki Prefecture	0	0	-
Tsushima City, Nagasaki Prefecture	0	0	-

Table 3: Municipalities that were river flood risk areas during Typhoon No. 16 in 2016 and the HAS scores.

where the HAS was calculated although there were no news reports, the Precision was determined to be 100%, the Recall was 57.1%, and the F-measure was 72.7%. Thus, it can be considered that this warning map has been confirmed as effective. However, there is a question as to why in Table 3, hazard warnings did not occur before the broadcast news in the areas listed from Oita-shi in Oita Prefecture and below. It seems that the reason is that with the free Twitter Streaming API, only 1% of the samples can be acquired. That is, even though 100 Tweets might exist, only 1 of these can actually be collected using the free API. In other words, if the premonitory symptom keywords remain buried in the other 99 Tweets, this can make a great difference in the HAS calculations.

Finally, in this research the evaluation experiment was carried out only for river overflows, but by using the premonitory symptom keywords shown in Table 4 and with the same procedures, the HAS can be also calculated for tsunami, landslides, and sudden downpours.

Japanese Disaster Warning Map (HAS heat map)

Figure 3 depicts the HAS for each area (municipality) visualized as a heat map (HAS heat map). In the HAS heat map, the process is done in the background (that is: 1. Tweets are gathered in real time→2. the HAS is calculated for the municipalities where the number of Tweets that include the premonitory symptom keywords exceeds 0.01% of the population of the municipality), and each time a HAS score is calculated it appears in the browser (i.e. the browser is renewed). In this way, when the user opens the browser to check the heat map, the latest information can always be displayed. Also, by checking at regular intervals, the information is regularly updated as appropriate while the user is viewing the browser. In Figure 3, the municipalities with a HAS score of 0 or over are shown in blue, those with a HAS score of 5 or over are in green, and those with a HAS Score of 10 and over are in shown in red. Each HAS Score visualized in Figure 3 (up to Takamatsu-shi in Kagawa Prefecture) represents an area where it is possible to give warnings before news broadcasts.

Related Works

Target disasters [source]	Premonitory symptom keywords
Landslides [9,10]	Ground rumbling, cracks, mountain rumbling, subsidence, muddy river, driftwood.
Tsunami [11]	Storm surge, ebbing tide
Sudden downpours [12,13]	Black clouds, thunder, cold wind, large raindrops, hail

Table 4: Keywords for the premonitory symptoms of landslides, tsunamis and sudden rainfalls.

Various research works has recently been carried out in the area of how to use SNS data to prevent disasters.

In 2012, Twitter Japan KK and Google KK organized The Great East Japan Earthquake Big Data Workshop [14,15]. At this workshop, the data generated in the week following the earthquake was presented and the participants reanalyzed the data in order to develop a variety of ideas and proposals. For example, proposals were made for using mobile phone GPS information to plot the congestion on a map in order to optimize evacuation guidance, or the mapping of Tweets with the latitude and longitude to estimate the extent of the damage from the numbers or times of the Tweets.

Sakamaki has attempted to build a system for understanding how the supply and demand of relief supplies is developing by using SNS data, such as Tweets, when a large-scale disaster occurs [16]. It is hoped that, by using this method, a solution can be found for the problem of mountains of relief supplies building up at depots after an earthquake, even though there are not enough relief materials at the shelters. Also, a specialist network research group is working, in response to the concern that in a large-scale disaster the existing network equipment monitoring will fail in a large-scale disaster, to understand the situation of the disaster from SNS data and to use this to detect any network system fault [17].

In addition, there is relevant research underway on the technology itself used for extracting events from social media such as Twitter. This includes: research discussing the technology required for extracting events effectively from a collection of short colloquial texts like Tweets [18], as well as research proposing technology to be used for extracting events according to their social importance [19]. Cerutti et al. [20] used

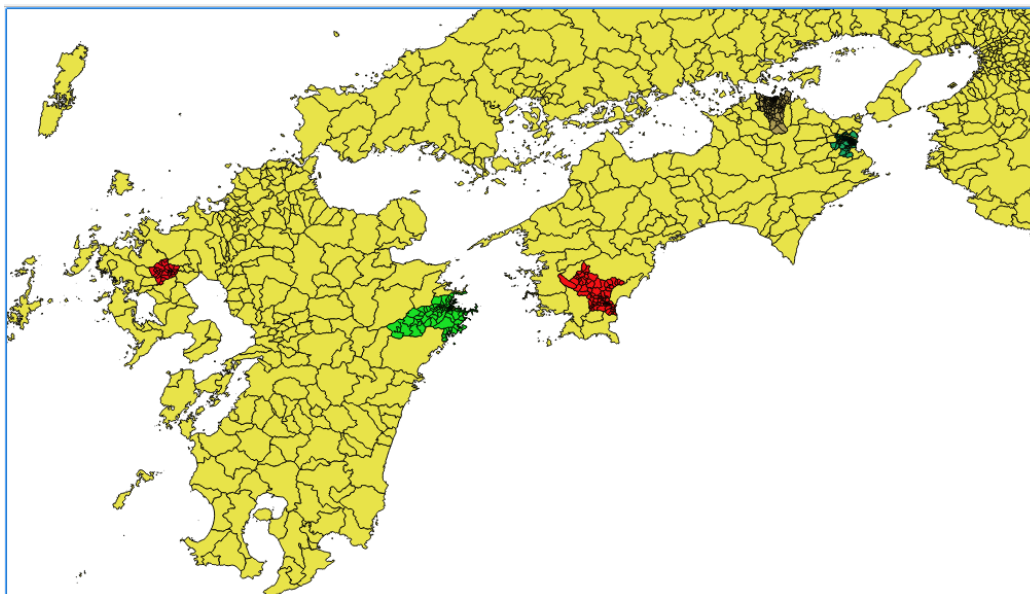


Figure 3: Japanese Disaster Warning Map (HAS heat map).

the keywords “flood” and “inundation” to pick out the areas suffering from damage and to visualize them, but their main purpose was to respond swiftly to disaster stricken areas after flooding had occurred, and not to give warnings before the news as is the case with the present research. Furthermore, Alsaedi et al. [21] extracted fires and traffic accident information efficiently using Tweeter hashtags, but their focus was on seeing if such phenomenon could be predicted, and it did not include visualization on a heat map to see the area with a wide extension.

Examples of research similar to this project include research by Ito et al. [22], work by the MLIT [23], and research by Sakai et al. [24] The research results related to the work by the MLIT on how to use Tweeted information to gather the premonitory symptoms of landslides was explained in Chapter of the Introduction. In their work, the disasters being considered just involve landslides, though they also suggest that it is possible to understand the level of urgency and state of the disaster from changes in the number of Tweets extracted according to the set of keywords or their content, by categorizing the Tweeted information relating to landslides and by setting the appropriate keywords. In contrast, the present research is targeted not only at landslides but also at river overflows, sudden downpours, and tsunamis. Sakai’s research also uses latitude and longitude as positioning information for the Tweets in performing an analysis, and refers to the Recall, Precision, and F-measure, but it differs from this research in that it does not extract the premonitory symptoms keywords from Tweets, but instead analyzes the keywords after the actual occurrence.

Finally, DISAANA [25] is a Web service used to automatically extract information about problems and trouble relating to disasters from Tweets occurring within a set of defined limits. It can then present possible answers to questions like “Where is heavy rain falling” in a list or a map, by analyzing the Tweets in real time based on the user’s search terms. However, it simply acquires and displays the Tweets related to disasters by connected them to the terms the user has searched for. In contrast, this research acquires the Tweets including premonitory symptom keywords in real time, then calculates the hazard coefficient from them, and visualizes these in a warning map. As this is a service that offers additional information by the further processing of the

acquired Tweets, it is believed to have a unique value.

Conclusion

In this research project, a disaster warning map was produced that warned of the scope of the danger before a disaster occurred by acquiring premonitory symptom keywords from Tweets in real time and by calculating a hazard coefficient from these Tweets. At present, the targeted disasters for these maps are landslides, tsunami, river overflows, and sudden downpours. In the evaluation experiment, the example of a river overflow was demonstrated. As a result, in 57% of the 21 cases where rivers overflowed or where the water surface exceeded the danger levels, it was shown that it was possible to make predictions before the time when news broadcasts were made.

Through further evaluation experiments in the future, it is expected that demonstrations involving disasters other than river overflows will be conducted, and through these demonstrations the premonitory symptom keywords that are used to gather the Tweets in real time will be refined.

References

1. Cabinet Office, Government of Japan: Disaster Management in Japan.
2. Cabinet Office, Government of Japan: Disaster Management Plan.
3. Ministry of Land, Infrastructure, Transport and Tourism, Government of Japan: Utilizing “Tweet information” to grasp the precursor phenomenon of landslide disasters.
4. Cabinet Office, Government of Japan Cabinet Office Report about the Damage Resulting from Typhoon No. 16 of 2016.
5. Twitter Developer Documentation “Places”.
6. Statistics Japan Website: Population Estimates 2015.
7. Shimosuwa City, Nagano prefecture Website.
8. <http://kotetsu.game-waza.net>.
9. Shizuoka Prefecture Website: Premonitory symptoms of landslide disaster.
10. Public Relation Office Government of Japan: Three points to protect yourself from landslide disasters.

11. National Geographic: Mechanism of the tsunami, premonitory symptoms, response to be taken.
12. Disaster Prevention System Institute Website.
13. Akiha ward, Niigata city Website: Refuge Guidebook of Akiha Ward.
14. Big Data Workshop of the 3.11 Great East Japan Earthquake.
15. Murakami K (2013) How to Utilize "Disaster Big Data": Prospects on the Use of Disaster Information in the Future. *The NHK Monthly Report on Broadcast Research* pp:2-25.
16. Sakamaki E (2012) Demand prediction system of relief supplies using Tweets data. *The Japan Society for Management Information, Autumn Nationwide Research Announcement Contest Proceeding* pp:319-322.
17. Maru C, Enoki M, Nakao A, Yamamoto S, Yamaguchi S, et al. (2016) Development of a Failure Detection System for Network Control using Collective Intelligence of Social Networking Services in Large Scale Disasters. In *Proc. of the 27th ACM Conference on Hypertext and Social Media (HT2016)* pp:267-272.
18. Atefeh F, Khreich W (2015) A Survey of Techniques for Event Detection in Twitter. *Comput. Intell.* 31:132-164.
19. Guille A, Favre C (2015) Event detection, tracking, and visualization in Twitter: A mention-anomaly-based approach. *Social Network Analysis and Mining* pp:5-18.
20. Cerutti V, Fuchs G, Andrienko G, Andrienko N, Ostermann F (2016) Identification of disaster-affected areas using exploratory visual analysis of georeferenced Tweets: Application to a flood event. *AGILE, Helsinki* pp:14-17.
21. Alsaedi N, Burnap P, Rana O (2017) Can We Predict a Riot? Disruptive Event Detection Using Twitter. *ACM Transactions on Internet Technology* 17:1 8.
22. Ito A, Kunitomo M (2015) Possibility to understand the situation of landslide disaster using tweets data. *64th Workshop of the Japan Society of Erosion Control Engineering* pp:132-133.
23. National Institute for Land and Infrastructure Management, Ministry of Land, Infrastructure and Transport: Approach to the detection of landslide premonitory symptoms by SNS data analysis.
24. Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes Twitter users: real-time event detection by social sensors. In *the Proceedings of the 19th International Conference on the World Wide Web (WWW '10)*, ACM, New York, USA pp:851-860.
25. <https://disaana.jp/rtime/search4pc.jsp>