E01

Statistical Text Analysis of Tweets Related to the 2015 Kinu River Flooding

○Yongxue SHI, Takahiro SAYAMA, Kaoru TAKARA

On September 10, 2015, due to heavy rainfall over three days, an overtopping floods occurred along the Kinu River in Japan around 6:00AM. At 12:50 PM on the same day, a levee breach also occurred at the downstream. This disaster left two people dead and several injured with enormous damages to homes and infrastructures. For effective countermeasures, real-time information on local flood situation is indispensable.

In the event of a disaster, quickly collect and share efficient damage information is very important for emergency response to the government as well as affected communities themselves.

In Japan, the national government collects disaster information at the cabinet information collection center 24 hours a day. National government providing local governments with necessary information about the hazard while local governments provide the national government agency with the damage information. National government has responsibility to oversee the while coordination process. And local municipalities have the direct responsibility to carry out emergency response operations, such as firefighting, rescue, ambulance serves (Cabinet office, Japan, 2011). Grasp the information on site timely and accurately played an important role in the municipalities' decision on issuing evacuation orders. Hyogo Framework for action also pointed out the importance of identify, assess and monitor disaster risks and enhance early warning for disaster management. In addition, flood hazard map as an important non-structure mitigation measure in disaster risk reduction also cannot be established without the efficient flood information.

As mentioned above, grasp the efficient flood information after a disaster occurs is the essential information to national and local governments as well as affected local communities themselves.

Based on literature review, researchers agreed that social media is useful and efficient for emergency responses. However, data collection and classification manually are massive work and consume a lot of time. Therefore, develop a system to collect the information and separate into several categories will increase the efficiency of decision-making.

The objectives of this research is to develop a way for text-mining which optimizes the needs of emergency response and maximize the use of social media for disaster risk reduction.

This research will mainly focuses on 1) tweets are automatically collected and classified into five groups –'inundation', 'rescue and relief' 'infrastructure damage', 'inundation with water depth' and others, by using text mining technology, 2) after classifying the information into the above five groups, then analyzing the information of each group and assuming to provide it to municipalities, fire fighter and polices, relief official volunteer organizations and residents.

In order to classify the social media information, text mining technology was applied. And in this research, SPSS text Analytics for survey and SPSS statistic were used. The detailed steps are as follows:

Access data: the supervised 170 tweets related to the 2015 Kinu River Flooding was collected.

Pre-processing: all the 170 tweets were saved into excel files.

Text processing:

1) SPSS text Analytics for survey did the

tokenization, filtering and lemmatization at the first step, as a result, the key word phrase were listed up.

2) These keywords were made as category input to SPSS Analytics again to output 170 tweets into structured data. The tweets are shown in 0,1 value related to keywords. For example, the tweet contain inundation and Kinu river will have value'1' under these two keywords and other keywords are marked with '0'.

3) As there are hundreds of key words were listed up from the previous step, in order to reduce the parameters for statistical analysis in next step, the same keywords were summarized into a common group manually. For instance, the keyword '浸水' '冠水' can be summarized into the same group 'inundation' , as a result of this part, the keywords are summarized into the 12 groups- 'inundation', 'water level', 'damage', 'traffic condition', 'time', ' location', 'levee', 'resuce'. 'relief', 'infrastructure damage' 'alert', 'Kinu River'. The tweets output also been modified with '0', '1' value based on the certain keywords.

4) The 170 original tweets were categorized into 5 groups as mentioned before, 'inundation', 'damage', 'rescue', 'inundation with water depth' and 'others'. The output from step 3 and the original groups will be used as supervised data for discriminant analysis. Discriminant analysis is a classification problem, where two or more groups or clusters or populations are known a prior or one or more observations are classified into one of the known population based on the measured characteristics (IBM). And with the function of discriminant analysis in SPSS statistic, trying to find the appropriate discriminant equation with the supervised data.

As a result, the canonical discriminant function coefficients are shown in Table 1, and the region map shown in Figure 1. With this canonical discriminant function coefficients, the raw data can get 4 values in four dimension, and refer to these four values, the raw data will be relocated to one of the 5 groups.

With a cross validation, the precise classification rate is 70%, which pointed out the canonical discriminant function coefficients are significant in this case.

正準判別関数係数

	関数			
	1	2	3	4
交通状況	-.293	-.197	.096	.905
時間	.527	-.217	-1.076	.853
地名	-.568	.430	-.105	-.200
浸水	-2.052	.622	-.019	1.281
鬼怒川	.186	.067	.801	.307
堤防	-.050	-.128	.696	-.591
救助	1.520	.304	-2.032	1.462
インフラ損害	1.567	-.558	1.080	.450
水位	.518	-.531	1.249	2.916
水深	-.608	-.110	.047	.086
警報	-.108	-.197	.345	.999
救援	1.527	7.612	2.035	.988
(定数)	1.094	-.688	-.347	-1.530

標準化されていない係数

Table 1 The canonical discriminant function coefficients

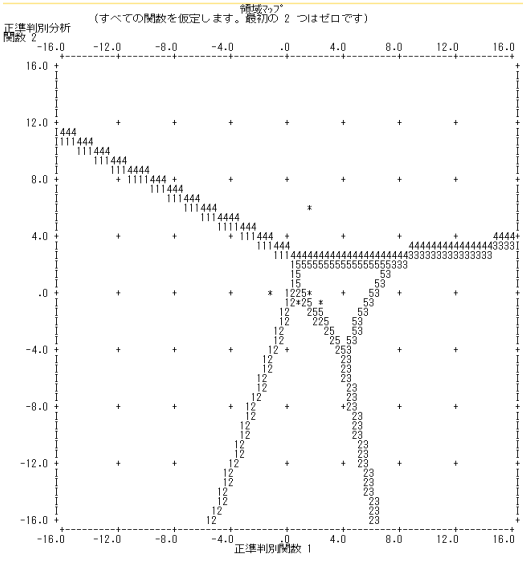

Fig.1 Region map

References:

Cabinet Office, Government of Japan. 2011. *Disaster Management in Japan* [PDF file]. Retrieved from http://www.bousai.go.jp/1info/pdf/saigaipanf_e.pdf

https://www.ibm.com