Social Media Analysis after the East Japan Great Earthquake in Thai and Japan

Takako Hashimoto
Chiba University of Commerce
Chiba, Japan
takak@cuc.ac.jp

Supavadee Aramvith
and Teeranoot Chauksuvanit
Chulalongkorn University
Bangkok, Thailand
supavadee.a@CHULA.AC.TH
and tchauksuvanit@gmail.com

Yukari Shirota
Gakushuin University
Tokyo, Japan
Yukari.shirota@gakushuin.ac.jp

Abstract—After the East Japan Great Earthquake occurred in Japan on 11th March 2011, many messages related to the earthquake were posted to social media. Especially, in Asian countries, a lot of messages about the earthquake were posted. Then, different topics such as damages by the earthquake and afflicted people were observed. These topics can be recognized as reactions to the earthquake. Therefore, exploring topics related to the earthquake on social media in Asian countries gains a rich insight into the Asian social context. The goal of our research is to analyze Asian people reactions to the East Japan Great Earthquake on social media using data mining technique. As the first step, this paper targets Thai language and conducts the preliminary approximation to investigate topics on Thai social media. We analyze how Thai people reacted to the earthquake and compare with reactions in Japan.

Keywords—Social Media, Data Mining, East Japan Great Earthquake, Thailand and Japan

I. INTRODUCTION

Social media in which individual users post their opinions and gradually build their consensus, is recognized as one of the important collaborations in today's information oriented society. After the topical problems like a disaster, people's behavior is influenced by this collaboration. Especially after the East Japan Great Earthquake occurred in Japan on 11th March 2011, many messages about the earthquake were posted to social media such as Twitter, Facebook, and so on, not only in Japan but

also all over the world. Particularly, in Asian countries, a lot of topics concerned about damages by the earthquake, afflicted people by the earthquake etc. were observed, and the topics triggered many kinds of supportive movements. Exploring topics related to the earthquake on social media in Asian countries gains a rich insight into the Asian social context.

We already proposed the graph based topic extraction method[1,2,3]. In our method, first, we crawl messages in social media and extract keywords using morphological technique. Next, we construct a snapshot document-term matrix at each time stamp. Then, we investigate topic transitions over time by forming network graphs from the matrix. Our method could show the time series structure transition by network graphs, so that we hypothesized that topics could be expressed by sub graphs. Because our method analyzes topic structures using keywords, once keywords were extracted, the method can be language-independent. As the first target, we selected Thai language. In Thai, social media is quite popular and after the East Japan Great Earthquake, a lot of people posted their messages about it on social media. We crawled social media messages related to the earthquake in Thai, extracted keywords from messages and analyzed topics as preliminary approximation. We also analyze how Thai people reacted to the earthquake and supported afflicted people by comparing reactions on social media between Thailand and Japan.

II.   THAI SOCIAL MEDIA ANALYSIS USING OUR METHOD

A.   Select Target Social Media in Thai

First of all, we had to select target social media from a lot of social media services in Thai. We examined 30 websites of Truehits.net Web Award 2011[4]. Among those, we selected the following 6 websites as social media candidates in Thai: www.kapook.com, www.sanook.com, www.pantip.com, www.dek-d.com, www.facebook.com, www.twitter.com.

B.   Preliminary Approximation for Social Media Message Crawling

To crawl messages from 6 social media candidates, we have to define the following 22 retrieval keywords (Tab. 1).

C.   Preliminary Approximation for Morphological Analysis for Thai Language

Unlike Japanese language, Thai language is a tonal language, which means that the same word can convey different meanings depending on the tone with which it is pronounced[5]. It is not easy to extract keywords from Thai language using morphological technique. To do word segmentation, we used the Swath program[6], which is word segmentation software for Thai. It is an open source program and can be used freely. As a result, a lot of errors occurred for work segmentation due to the characteristics of Thai language.

Table I. Retrieval keywords

#	Keywords in Thai	Translation in English
1	สึนามิ 2554	Tsunami 2011
2	สึนามิ and 2554	Tsunami and 2011
3	สึนามิญี่ปุ่น 2554	Tsunami in Japan/ Japan tsunami 2011
4	สึนามิญี่ปุ่น and 2554	Tsunami in Japan and 2011
5	สึนามิญี่ปุ่น 2554 not สึนามิ	Tsunami in Japan and 2011
6	สึนามิญี่ปุ่น 2554 or แผ่นดินไหวญี่ปุ่น 2554	Tsunami in Japan 2011 or Japanese earthquake 2011/ Earthquake in Japan 2011
7	สึนามิ 2011	Tsunami 2011
8	สึนามิ and 2011	Tsunami and 2011
9	สึนามิญี่ปุ่น 2011	Tsunami in Japan/ Japan tsunami 2011
10	สึนามิญี่ปุ่น and 2011	Tsunami in Japan and 2011
11	สึนามิญี่ปุ่น 2011 not สึนามิ	Tsunami in Japan and 2011
12	สึนามิญี่ปุ่น 2011 or แผ่นดินไหวญี่ปุ่น 2011	Tsunami in Japan 2011 or Japanese earthquake 2011/ Earthquake in Japan 2011
13	แผ่นดินไหวที่ญี่ปุ่นปี 2554	Earthquake in Japan 2011
14	แผ่นดินไหวที่ญี่ปุ่น and ปี2554	Earthquake in Japan and 2011
15	แผ่นดินไหวญี่ปุ่น 2554 not แผ่นดินไหว	Earthquake in Japan and 2011
16	แผ่นดินไหวที่ญี่ปุ่นปี 2011	Earthquake in Japan 2011
17	แผ่นดินไหวที่ญี่ปุ่น and ปี2011	Earthquake in Japan and 2011
18	สารกัมมันตรังสี	Radioactive elements
19	8.9ริกเตอร์	8.9 Richter
20	8.9 ริกเตอร์	8.9 Richter
21	สึนามิ	Tsunami
22	แผ่นดินไหว	Earthquake

We've corrected errors manually for keywords, and finally, 108 keywords were extracted from the topic titled "Japan earthquake 7.1 Richter and tsunami warnings canceled" on www.Kapook.com. The topic consisted of 1508 messages from 11 March 2011 to 15 April 2011. In 108 keywords, there are words related to the nuclear such as "radio activity," "nuclear power plant," and "nuclear reactor." There are also supportive words such as "worry," "mindful," "Don't give up," "help," and "love," and words about earthquake damage such as "natural disaster," "tsunami," and "shake." These words show that there are topics for the nuclear plant accident, afflicted people and the severe earthquake.
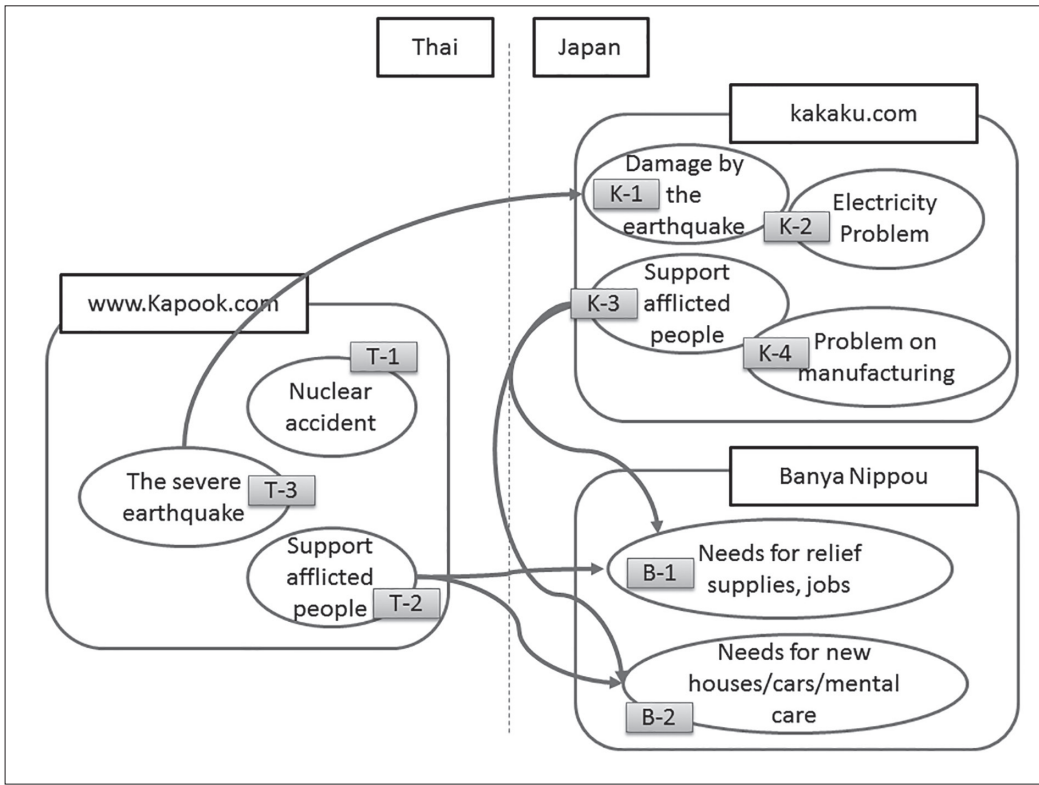
Fig. 1.   Reaction in social media after the earthquake between Thai and Japan

III.   COMPARISON OF REACTION IN SOCIAL MEDIA AFTER THE EAST JAPAN GREAT EARTHQUAKE BETWEEN THAILAND AND JAPAN

This section discusses the comparison of reactions in social media after the East Japan Great Earthquake between Thailand and Japan. In our previous work, we analyzed topics in the kakaku.com[7] and Banya Nippou[8] in Japan using our method. The relationships between the above 3 social media in Thai and Japan are shown in Fig. 1. K-3 and T-2 are topics for supporting afflicted people so that they can be recognized reactions for afflicted people needs (B-1 and B-2). K-1 and T-3 are similar topics, so that it shows that even in Thai, people also discussed damage by the earthquake. K-2 is about the electricity problem that happened immediately after the earthquake in Japan. It was also important topic at that time. Regarding T-1 about the nuclear accident, we could find it in Thai for now. But we expect it be able to be found in Twitter in Japan, because Twitter is more lively and casual social media. We plan to investigate Twitter as well.

IV.   CONCLUSION

This paper described the preliminary approximation to apply our social media analysis method to Thai language. We discussed morphological analysis errors caused by characteristics of Thai language. Based on our approximation, we also analyzed how Thai people reacted to the earthquake by comparing reactions on social media between in Thailand and in Japan.

ACKNOWLEDGMENT

References

[1] Hashimoto, T., Kuboyama, T., Chakraborty, B., Shirota, Y., "Discovering Topic Transition about the East Japan Great Earthquake in Dynamic Social Media," GHTC 2012, pp.259-264, 2012.

[2] Higuchi, S., Hashimoto, T., Kuboyama, T., Hirata, K., "Exploring Social Context from Buzz Marketing Site – Community Mapping Based on Tree Edit Distance –," Proc. of PerCol 2013 (Fourth International Workshop on Pervasive Collaboration and Social Networking), pp.187-192, 2013.

[3] Hashimoto, T., Chakraborty, B., Kuboyama, T., Shirota, Y.,"Temporal Awareness of Needs after East Japan Great Earthquake using Latent Semantic Analysis," Proc. of the 23rd European-Japanese Conference on Information Modelling and Knowledge Bases (EJC2013), pp.214-226, 2013

[4] Thai Internet Innovation Research Institute, "Top Website Award Truehits.net Web Award 2011," 2012, http://truehits.net/ [2012,July 24].

[5] Basic Introduction to Thai Language, http://wws.peacecorps.gov/wws/multimedia/language/transcripts/TH_Thai_Language_Lessons.pdf

[6] Software: SWATH – Thai Word Segmentation, http://www.cs.cmu.edu/˜paisarn/software.html

[7] Banya Nippo provided by SAVE IWATE (Non-profit organization), http://sviwate.wordpress.com/in-english/

[8] kakaku.com, http://kakaku.com/