Japan Geoscience Union Meeting 2012
(May 20-25 2012 at Makuhari, Chiba, Japan)

MTT38-05 Room:301A Time:May 21 16:30-16:45

Gathering, analysis, and visualization of meteorological phenomena using social media

KITAMOTO, Asanobu[1*]

[1]National Insitute of Informatics, [2]JST

Meteorological phenomena have influence on all the people, so it is one of the most popular topics in daily conversations on social media. On the other hand, social media can also be used as a place for reporting emergency information instantly during meteorological hazards such as typhoons, and there has been strong expectation that social media will evolve into an effective platform for collecting and sharing on-site reports for emergency response. Hence this paper focuses on meteorological phenomena and social media, and discusses how we can gather, analyze, and visualize social media based on our experiences.

This paper mainly deals with the case study of four systems, namely ”Typhoon Front” (http://front.eye.tc/), ”Twiphoon” (http://twiphoon.eye.tc/), ”Typhoon Now!” (http://typhoonnow.eye.tc/), and ”Futtekitter” (http://agora.ex.nii.ac.jp/futtekitter/). We started an experiment of collecting typhoon information from weblogs in 2004, and from Twitter in 2009. These services are based on a ”call-based gathering method” that requires user’s actions to report events. This method has an advantage of controlling the format and quality of user’s reporting, but it has limitation in the amount of information collected due to the necessity of user’s active involvement. Hence we employed a ”patrol-based gathering method” for ”Futtekitter” started in 2012, and build a system that actively gathers information about events without explicit involvement of users. This system increased the amount of information we can collect to the level of several thousand tweets per hour, for instance, during the event of snowfall in Kanto Region.

The problem of a patrol-based gathering method, however, is that we cannot expect users to assign well-organized metadata. This problem is especially apparent in the problem of named entities. For example, which place and what kind of events a tweet is referring to? This kind of metadata is sometimes given by users in the form of hashtags on Twitter, but it is generally difficult to control because there is no single standard for this kind of metadata. To solve this problem, we need to apply natural language processing (NLP) to natural language text for analyzing content, extracting named entities, and integrating them. To realize this process for placenames, we are now developing a geocoding tool, GeoNLP (http://agora.ex.nii.ac.jp/GeoNLP/), which can extract place names from text and resolve them to a unique location. This tool still has a problem of accuracy, but the automatic mapping of text is now easier than before.

We built ”futtekitter” with functionalities addressed above, and applied it to the gathering, analyzing, and visualizing tweets on rain and snow. By analyzing tweets, to what spatial and temporal resolution we can collect information about rain and snow from social media, and to what degree we can rely on uncertain information? We discuss the result of analysis based on this motivation.

Keywords: Meteorological phenomena, Social media, Twitter, Named entity, Geocoding, Natural language processing