Information Current in Twitter: Which Brings Hot Events to the World

Peilei Liu, Jintao Tang, Ting Wang
College of Computer, National University of Defense Technology
410073 Changsha, Hunan, China
{plliu, tangjintao, tingwang}@nudt.edu.cn

ABSTRACT

In this paper we investigate information propagation in Twitter from the geographical view on the global scale. An information propagation phenomenon what we call "information current" has been discovered. According to this phenomenon, we propose a hypothesis that changes of information flows may be related to real-time events. Through analysis of retweets, we show that our hypothesis is supported by experiment results. Moreover, it is discovered that the retweet texts are more effective than common tweet texts for real-time event detection. This means that Twitter could be a good filter of texts for event detection.

Categories and Subject Descriptors

H.2.8 [Database Management]: Database Applications – Data Mining; H.3.5 [Information Storage and Retrieval]: On-line Information Services – Web-based services

Keywords

Twitter, information propagation, geography, event detection

1. INTRODUCTION

Twitter is a popular online micro-blog service which allows users to send short texts and share information through the retweet activity. Most existing researches of information propagation in Twitter concentrate on analyzing the influence of network structure on information propagation [1,2], while most researches of real-time event detection concentrate on analyzing common tweet texts [3-5]. In this paper we present the study of retweet behavior from a geographical viewpoint and its potential usefulness for real-time event detection.

An important characteristic of micro-blog services is their real-time nature. Due to this characteristic, Twitter has been viewed as social sensors in existing works. Sakaki et al. proposed an algorithm to monitor tweets and detect earthquakes [3]. Bollen et al. analyzed the mood of tweets to predict the stock market [4]. In our opinion, however, Twitter could also be a good filter for event detection. Intuitively, users tend to forward the tweets related to some important real-time events, which indicates that retweets should be more effective than common tweets for event detection.

WWW 2013 Companion, May 13-17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2038-2/13/05.

According to this idea, we investigate retweet behavior from the geographical view. An interesting phenomenon what we called "information current" is discovered. According to this, we propose a hypothesis that the changes of information flows in information current may be related to real-time hot events. Some empirical rules are investigated through the analysis of data. We find that the information current brings hot events, and noises are turned down during the retweeting process. Therefore, retweet texts are more valuable for real-time event detection than common tweet texts.

2. ANALYSIS AND RESULTS

We crawled approximately 50 million tweets using the Twitter API in August 2011. In this dataset, user's geographical positions and the time tweets published were extracted. There are several attributes related to geographical positions. Although coordinate is the best one, tweets with coordinate are very few. Registered locations are free-texts, so the synonym problem is a difficulty. We choose time-zones as the indicator of users' geographical positions finally, because they are chosen from a normalized time-zones list. In addition, a time-zone, which has different names in different countries, can represent a country or a part of it.

2.1 Information Current Phenomenon

Representative time-zones, such as London, Eastern Time (US & Canada) are extracted following two principles: 1) Choosing thsose publishing the most tweets. 2) Choosing the most representative ones in every continent. And we use the number of retweets between time-zones as the indicator of the information flow strength.

Generally speaking, the strength of an information flow from one position to another is unequal to the reverse flow. These two flows can be integrated into a single unidirectional flow. On the global scale, all these unidirectional flows between time-zones together will generate the information current phenomenon, like the ocean current. The main factor influencing these global information flows are geographical positions. They are relatively regular and steady. In general, more information flows into developing countries from developed countries. However, through analysis of the dataset, we find that some changes of the flows' strength occur every day. And this may be due to real-time events, according to our previous hypothesis. So we will have a different global information current graph every day.

2.2 Information Current and Real-time Event

According to our previous hypothesis, changes of information flows may be caused by real-time events. Specifically, hot events shift the directions of users' attention, which then affect the number of retweets. Therefore we can detect real-time events through monitoring the changes of the information current graph. We analyze information flows between main time-zones in the dataset, and find that the strongest flows approximately fit the headline news in news sites.

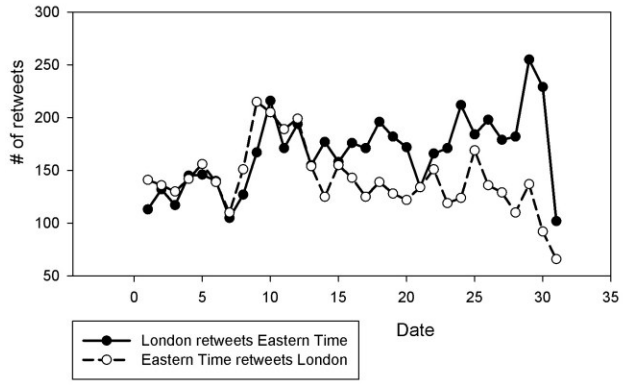

Figure 1. Flows between London and Eastern Time.

Table 1. Events related to UK and US in August 2011

Date	Event
6	London Riots
18	Israel Bombing Gaza Strip
21	Syria War
27	America Irene Hurricane

As an instance, the bidirectional information flows between London and Eastern Time (US & Canada) have been presented in Fig. 1. The flows become stronger in some specific dates. For example, from 7 to 10 America retweet England more than usual; while from 27 to 30 England also retweet America abnormally. Headline news related to UK and US are collected and shown in Table 1. Overall, in the very day of an event, the corresponding information flow hits the bottom, and then becomes stronger. Two or three days later, it will reach to the peak. However, this doesn't mean any time lags, because the date, when the information flow hitting the bottom could be viewed as an early warning.

2.3 Real-time Events Detection

We then investigate the retweet texts for hot event detection. We collect the bidirectional retweets between London and Eastern Time in specific dates and generate tag clouds respectively. Parts of results are shown in Fig. 2: a is generated from 34k retweet texts from London to Eastern Time in Aug 9, while c is from 24k retweets from Eastern Time to London in Aug 28. As a comparison, we collect tweets published in London and Eastern Time in the corresponding date and generate tag clouds. In Fig. 2, b is generated from 4500k tweets published in London in Aug 9, and d is from 6600k tweets published in East Time in Aug 28.

From a and b in Fig. 2, you can discover that America retweet England a lot in Aug 9 because of the London riots. From c and d, you can discover that England pay more attention to America for the Irene Hurricane. Comparing a and b, it can be found that they are similar, but texts used in a is much fewer. Comparing c and d, it can be found that c is actually better, although using much fewer texts. There exist some noises in d, such as "lol", "don" "love" and "thanks", while these noises have been taken down in c. We have gotten similar results in many experiments on different dataset. This indicates that Twitter can be viewed as a noise filter of texts for event detection.

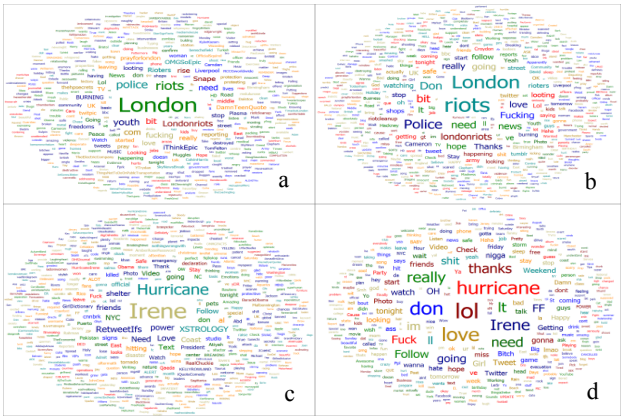

Figure 2. Tag clouds for event detection

3. CONCLUTION

We investigate the retweet behavior from the geographical view and present the information current phenomenon in this paper. Changes of information flows are then analyzed for real-time event detection. In addition, we demonstrate that Twitter could be a filter of texts for event detection. A limit of this work is that the strengths of information flows across different countries are influenced by the differences of languages. However, the ideas shown in this paper can also be used to analyze the information current between cities within a country, only if the corresponding geographical information is available.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 61202337, 61170156, and 60933005).

4. REFERENCES

[1] E., Bakshy, I., Rosenn, C., Marlow, and L., Adamic. The role of social networks in information diffusion. In *Proc. WWW*, 2012.

[2] M., Kim, L., Xie, and P., Christen. Event diffusion patterns in social media. In *Proc. ICWSM*, 2012.

[3] T., Sakaki., M., Okazaki., and Y., Matsuo. 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proc. WWW*, 2010.

[4] J., Bollen, H., Mao, and X., Zeng. Twitter mood predicts the stock market. *J. Computational Science* 2(1) (2011), 1–8.

[5] Y., Hu, A., John, D.D., Seligmann, and F., Wang. What were the tweets about? Topical associations between public events and Twitter feeds. In *Proc. ICWSM*, 2012.