Measurement of Burst Topic in Microblog

Guozhong Dong¹, Xin Zou²*, Wei Wang¹, Yaxue Hu¹, Guowei Shen¹, Korawit Orkphol¹, Wu Yang¹

¹Information Security Research Center, Harbin Engineering University
Harbin, Heilongjiang Province, China 150001
²National Computer Network Emergency Response Technical Team/Coordination Center, Beijing, China 100029
dongguozhong@hrbeu.edu.cn, zouxin@cert.org.cn, w_wei@hrbeu.edu.cn, huyaxue@hrbeu.edu.cn, shenguowei@hrbeu.edu.cn, korawit.orkphol@gmail.com, yangwu@hrbeu.edu.cn

Abstract. Microblog provides the first communication platform for burst event due to the immediacy and interactivity of microblog. In this paper, we research on user-oriented and message-oriented measurements of burst topic in Sina microblog. The measurements and analysis on large-scale Sina microblog data set show that our proposed measurement method can measure the characteristics of user and message propagation in burst topic. The measurement results in this paper can contribute to better research of relevant issues on burst topic and ensure the well-developed of microblog.

Keywords: Sina microblog, burst topic, user-oriented measurement, message-oriented measurement

1 Introduction

Microblog is a communication and user-generated platform which makes everybody be producer, communicator and commentator. Since microblog is great source for getting access to information much faster than traditional sources of media, more and more organizations and public figures post and spread information through microblog. Because of immediacy and interactivity of microblog, microblog provides the first communication platform for burst event as soon as event occurs. For example, oil explosion accident in Qingdao, microblog is the earliest news source. The reports and discussions for burst event are benefit to crisis response and situational awareness. But because of spreading messages almost instantly and spotty user quality, it will speed up the propagation velocity and have a negative effect on controlling network public opinion when burst topics are spread by malicious users within a short time. All these issues have brought challenges to study on burst topic in microblog including burst topic detection, mining key users of burst topic, control and prediction of topic diffusion. Faced with these challenges, we research on user-oriented and message-oriented measurements of burst topic in microblog. The measurement results in this paper contribute to better research of relevant issues on burst topic and ensure the well-developed of microblog. Aditi et al.[1] analyze the credibility of information...
in tweets corresponding to fourteen high impact news events and identify the important content and sourced based features to predict the credibility of information in a tweet. Yan et al.[2] prove that the small-world characteristic of microblog social network and the degree distributions of users are power-law. A social network based on human dynamics model is proposed based on their empirical analysis. Fan et al.[3] measure the topological characteristics and user behavior patterns in Sina microblog which are helpful for monitoring and controlling the microblog. Liu et al.[4] conduct the topic-oriented research on the measurement from many aspects such as features of the content, the network topology and user behavior. The measurement indicators and results can be effectively applied in a topic-generated network. To the best of our knowledge, together with the recent studies on social network, we are among the first to make static and dynamic measurement of burst topic in Sina microblog.

2 Measurements and analysis of burst topic

Considering the characteristic of real-time and huge data, we developed web crawler to collect data set of burst topics. 16 volunteers were interested in our work and participated directly in the data collection process. Before crawling data, volunteers selected 28 burst topics in Sina microblog and labeled hashtags for each burst topic according to the keywords of burst topics. These labeled hashtags were used as keywords to collect messages and user information of burst topic. In order to better analysis and measurement, we selected the most representative burst topic (Jaycee Chan taking drug) as the measurement object.

2.1 User-oriented measurement and analysis of burst topic

The number of user’s followers can reflect user influence to some extent. The distribution of user’s followers is shown in Figure 1. As shown in Figure 1, about 85% of users have less than 3,000 followers and 55% of users have 300-3,000 followers. Only a small percentage of users have a large number of followers.

![Fig. 1. The distribution of users’ followers](image)

Copyright © 2014 SERSC
In order to measure the time distribution of post behavior which reflects massive users’ behavior characteristic in burst topic, we set time window to be 1 hour and the time range to be 72 hours from the day that burst event happened. The time distribution of message in burst topic is shown in Figure 2. As shown in Figure 2, the post behavior not only conforms to bedtime and rising time, but also the evolution of burst topics. There are few messages about burst topics between 1 a.m and 7 a.m. Through the analysis of message content, we can find that users’ post behaviors change with the development and evolution of burst topics at other times.

![Fig. 2. The time distribution of message in burst topic](image)

### 2.2 Message-oriented measurement and analysis of burst topic

In the process of topic propagation, some messages have a significant increase during a certain time interval which is the major cause of burst topic. Six representative spikes of online media are shown in [5]. We can detect burst messages based on message’s rise-and-fall patterns. To model message’s burst attribute, messages in message forwarding tree will be divided into different time windows based on the post time of message to obtain the time sequence of message. The burst detection approach based on technical analysis indicators such as EMA, MACD, and MACD histogram is defined in [6]. A burst is defined as a time interval in which the MACD Histogram value is greater than burst threshold. The strength of the burst at a specific timestamp is the MACD Histogram value of that timestamp. MACD Histogram values of different time windows are shown in Figure 3. The message has burst attribute between the eighth and twelfth time window when burst threshold is set to be 30. Through analysis of users that posted burst messages, we found that burst messages are the key nodes in the development of burst topic. Detecting burst message in real-time can contribute to detect burst topic and key users in burst topic.
Fig. 3. MACD Histogram values of different time windows

Through the analysis of influential users’ message forwarding tree as shown in Figure 4, the propagation topology of forwarding tree can be divided into two categories: star propagation and tree propagation. The sources of star propagation are almost the users that post original messages and trendsetters in burst topic. Most of users in star propagation structure are source’s followers and mainly in the first level of propagation tree. Tree propagation structure may consist of more than one influential user. Each propagation structure of influential user can be seen as star propagation. Based on the analysis of the propagation topology, the width and depth of message propagation were measured. We found that the depth of forwarding tree is small and almost all of messages’ depths are below 6. However, the widths of popular messages are large, especially compared with the depths.

Fig. 4  The propagation topology of forwarding tree

3 Conclusion

In this paper, we have addressed the problem of measurement and analysis of burst topic in microblog. An effective measurement approach based on user entity and message entity is presented. The measurements and analysis on large-scale Sina
microblog data set show that our proposed measurement method can contribute to better research relevant issues on burst topic.

Acknowledgement. The authors would like to thank the reviewers for suggesting many ways to improve the paper. This work was partially supported by the National High Technology Research and Development Program of China(no. 2012AA012802), the National Natural Science Foundation of China(no. 61170242, no. 61101140, no. 61272537) and the Fundamental Research Funds for the Central Universities(no. HEUCF100611).

References

3. 樊鹏翼, 王晖, 姜志宏, 李沛. 微博网络测量研究. 计算机研究与发展. 49, 4 (2012)
4. 刘玮, 王丽宏, 李锐光. 面向话题的微博网络测量研究. 通信学报. 34, 11 (2012)