Discovering burst patterns of burst topic in twitter

Guozhong Dong[a], Wu Yang[a,*], Feida Zhu[b], Wei Wang[a]

[a] *Information Security Research Center, Harbin Engineering University, Harbin, 150001, China*
[b] *Singapore Management University, Singapore*

ABSTRACT

Twitter has become one of largest social networks for users to broadcast burst topics. There have been many studies on how to detect burst topics. However, mining burst patterns in burst topics has not been solved by the existing works. In this paper, we investigate the problem of mining burst patterns of burst topic in Twitter. A burst topic user graph model is proposed, which can represent the topology structure of burst topic propagation across a large number of Twitter users. Based on the model, hierarchical clustering is applied to cluster burst topics and reveal burst patterns from the macro perspective. Frequent sub-graph mining is used to discover the information flow patterns of burst topic from the micro perspective. Experimental results show that several interesting burst patterns are discovered, which can reveal different burst topic clusters and frequent information flows of burst topic.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

With the development of web 2.0, social media services, such as Twitter, emerge and quickly become popular. Different from traditional news media, Twitter allows users to broadcast short textual messages and express opinions using web-based or mobile-based platforms. When breaking news or events occur, people can post tweets about breaking news and share with friends. Due to large number of people participating in conversation and discussion, some tweets may become hot messages and the source of burst topics. Fig. 1 illustrates the user engagement time series of two burst topics detected by CLEar(Clairaudient Ear) system,[1] in which the arrow denotes each detecting time of burst topic. For example, Fig. 1(a) shows the burst topic that prominent Chinese human rights lawyer Pu Zhiqiang was set to stand trial in Beijing court. The topic was caused after Twitter user (BBCNewsAsia) post a tweet about the event. As shown in Fig. 1(a), the topic had one burst and was detected once by CLEar system in its lifecycle. Fig. 1(b) shows the burst topic about promotion activity over the Christmas period, which had more than one burst in its lifecycle. The different burst patterns raise a question of immense practical value: Can we leverage burst topics detected by CLEar system to discover burst patterns of burst topic in Twitter?

Unfortunately, mining burst patterns in burst topic has not been solved by the existing works. Most prior research works [1-14] focus on detecting burst topics in social media, instead of mining burst patterns in our work. Shen et al. [14] analyze the burst pattern of burst keyword, which can influence the accuracy of burst topics detection. In conclusion, burst pattern of burst topic is an important factor in the studies on burst topic.

---

* Corresponding author.
  E-mail addresses: dongguozhong@hrbeu.edu.cn (G. Dong), yangwu@hrbeu.edu.cn (W. Yang), fdzhu@smu.edu.sg (F. Zhu), w_wei@hrbeu.edu.cn (W. Wang).
  [1] http://research.pinnacle.smu.edu.sg/clear/.

(a) The burst topic about Chinese human rights lawyer Pu Zhiqiang

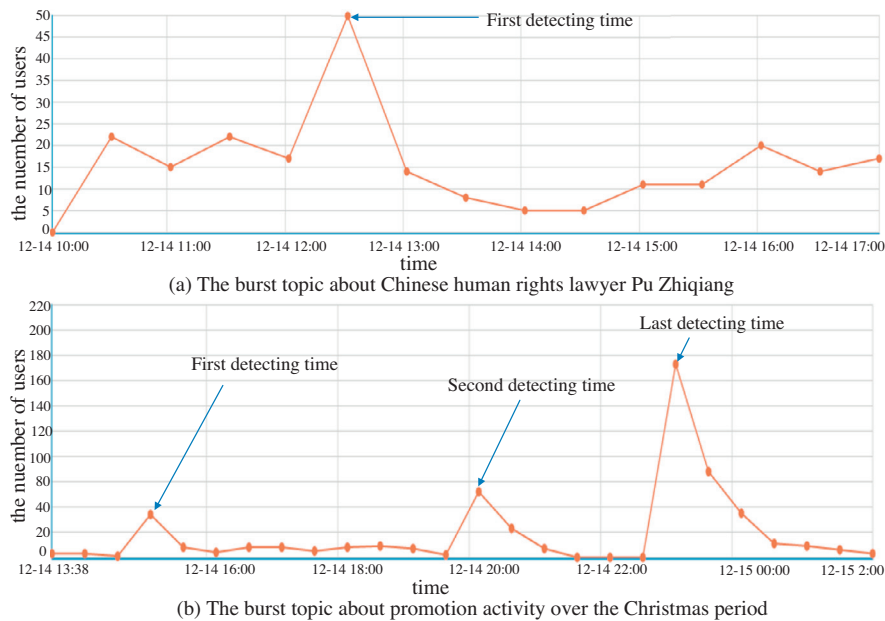(b) The burst topic about promotion activity over the Christmas period

Fig. 1. Example of user engagement time series of burst topic.

In this paper, we investigate the problem of mining burst patterns of burst topic in Twitter. To solve this problem, we discover burst patterns from both macro and micro perspectives by leveraging a burst topic user graph model. To summarize, the contributions of our work are listed as follows:

(1) We propose a burst topic user graph model which can represent the topology structure of burst topic propagation across a large number of Twitter users. In the burst topic user graph, nodes represent the burst topic users and edges represent the follower/followee relationship between users.
(2) Hierarchical clustering is applied to cluster burst topics and reveal burst patterns from the macro perspective. Combined with extracted 12 topic features, four distinct clusters of burst topic are discovered, which correspond to different burst patterns.
(3) Frequent sub-graph mining is used to discover information flow patterns of burst topic from the micro perspective. Based on the frequent sub-graph mining, several information flow patterns are extracted, which can be applied in several potential applications.

The rest of the paper is organized as follows. Section 2 reviews the related work. Section 3 presents the definition and construction of burst topic user graph. Macro and micro burst pattern mining are presented in Section 4. Section 5 describes the experimental results and findings. Finally, we conclude our work in Section 6.

2. Related work

The study of burst topic in social media and data mining techniques [1–19] have been studied in the last decade. As there are numerous research works focusing on it, here we introduce the ones most related to our work.

Prasadet et al. [1] propose a framework to detect emerging topics through the use of dictionary learning. They determine novel documents in the stream and subsequently identify cluster structure among the novel documents. Agarwal et al. [2] model emerging events detection problem as discovering dense clusters in highly dynamic graphs and exploit short-cycle graph property to find dense clusters efficiently in microblog streams. Alvanaki et al. [3] present the "en Blogue" system for emergent topic detection. En Blogue keeps track of sudden changes in tag correlations and presents tag pairs as emergent topics. Mathioudakiset et al. [4] identify burst keywords and group burst keywords into topics based on their co-occurrences. Cataldi et al. [5] formalize the keyword life cycle leveraging a novel aging theory intended to mine burst keywords and detect burst topics through keyword-based topic graph. Nguyen et al. [6] introduce a novel concept of sentiment burst and employ a stochastic model for detecting bursts in text streams. Takahashi et al. [7] apply a recently proposed change-point detection technique based on Sequentially Discounting Normalized Maximum Likelihood (SDNML) coding to detect abnormal messages and detect the emergence of a new topic from the anomaly measured through the model. Cui et al. [8] study some event-related properties of hashtags, including temporal trends, authorships and pattern of texts. Based on event-related properties of hashtags, they examine the popular hashtags to discover breaking events. Li et al. [9] propose "Twevent" system to detect events in twitter stream which can distinguish the realistic events from the noisy ones. Lee