Post2Event: Extracting Key Events from Microblogs

Chongwei Wang [a], Xujian Zhao [a*], Peiquan Jin [b], Hui Zhang [a], Chunming Yang [a], Bo Li [a]

[a] *Southwest University of Science and Technology, Mianyang 621000, China*

[b] *University of Science and Technology of China, Hefei 230027, China*

* jasonzhaoxj@gmail.com

Abstract— **This paper demonstrates a prototype system called Post2Event that aims to extract key events from microblogs. While many events are hidden in microblogs, people may only care about those critical events, which are named key events in Post2Event. Specially, we propose to model the topic-related significance of an event and integrate the influence with the temporal characteristics of the event to measure the event's importance. We briefly present the architecture and technical details of Post2Event. Then, we report the comparative results of Post2Event on a real dataset. Finally, we demonstrate the running process of the system.**

Index Terms—**Event extraction, Social influence, Temporal distribution, Microblog**

I. INTRODUCTION

Nowadays, social events occur every day and spread globally, forming a vast information source of social events. Meanwhile, news hotspots propagate through social networks and develop time-varying topics. This paper presents a prototype system called Post2Event to detect the key events in time-varying topics. Such a system can help users quickly understand the development of events and help decision-makers analyze the public situation on microblogging platforms.

Previous solutions to event extraction from microblogs relied on textual similarity, e.g., the TF-IDF similarity. However, the TF-IDF model always represents an event by some keywords and can only reflect the word-level textual importance of an event. Other graph-based approaches proposed to model events as graphs and use dominating set algorithm [1] or clustering coefficient information [2] of graph nodes to extract key events, but they failed to capture the time dimension of events, i.e., the importance of event may vary with time and evolution of topics.

Differing from existing approaches, We propose a new way to measure the importance of events. Specially, we propose to model the topic-related social influence of an event and integrate the influence with the temporal characteristics of the event to measure the event's importance. Following this idea, we develop the Post2Event system and demonstrate that it has higher effectiveness of key-event detection than existing methods.

II. DESIGN AND IMPLEMENTATION OF POST2EVENT

The architecture of Post2Event is shown in Fig. 1. It consists of three modules. The *Preprocessing* module is to clean data and generate candidate event set. The *Social Influence* module is to model the significance of events, and the *Temporal Distribution* module is to model the temporal distribution of events and extract key events.
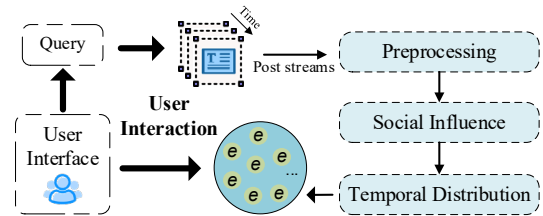
Fig. 1. The architecture of Post2Event

The input of the system is a set of microblogs, $P = \{p_1, p_2, \dots, p_i, \dots, p_n\}$, and each p_i is a timestamped microblog. The output of the system is a time-ordered list of key events.

A. Social Influence

Key events need to be able to represent the important information of topics on different timestamps fully. For social networks, opinion leaders are more influential to information diffusion because they usually deliver more critical information than ordinary users. To this end, microblogs posted by opinion leaders are more likely to be key events. Therefore, the forwarding number, commenting number, and the total number of likes are used to calculate an event's representativeness. Specifically, the social influence of an event can be represented by Eq. 1.

$$SI(e_i) = \log(\alpha \cdot f_n + \beta \cdot c_n + \gamma \cdot l_n + \varepsilon) \tag{1}$$

Here, the numbers of forwards, comments, and likes are denoted as f_n, c_n, and l_n, respectively. The symbol α, β, and γ are the corresponding weights, and ε refers to the Euler number.

B. Temporal Distribution

Time is one of the essential characteristics of events. By modeling the time characteristics of events, it can effectively reflect the evolution of events with time. For events happening simultaneously, users may have limited attention, which means that users usually focus on events with high social influence. Temporal expressions in an event are regularized by the normalizing algorithm [3] and we consider the social influence of an event in different timestamps to extract key events.

For an event sequence $E = \{e_1, e_2, \dots, e_i, \dots, e_n\}$, we record the events at each timestamp, obtaining the recording series of events $E_t = \{T_1 : [\cdots, e_{i-1}, e_i, e_{i+1} \cdots], T_2 : [\cdots, e_{j-1}, e_j, e_{j+1} \cdots], \cdots\}$. The social influence $IW(T_i)$ in T_i is defined as:

$$IW(T_i) = \sum_{e_1}^{e_m} SI(e_i) \tag{2}$$

The number of events extracted at timestamp T_i is defined as:

$$N(T_i) = \left\lfloor \frac{IW(T_i) - Min\_IW}{Max\_IW - Min\_IW} * n \right\rfloor \quad (3)$$

Here, $IW(T_i)$ is the sum of social influence in T_i. Max_IW and Min_IW represent the maximum and minimum of social influence at each timestamp, respectively. n is a constant that indicates the maximum value of key events extracted from each timestamp. And the outermost bracket denotes a rounding down function.

III. Performance of Post2Event

We evaluate the performance of Post2Event a real dataset collected from Sina Weibo (https://weibo.com). Three baseline algorithms are selected for comparison:

(1) TF-IDF [4]. It selects the events with a high TF-IDF score as key events.

(2) MWDS [1]. It constructs a graph representing the similarity among events and uses the Minimum-Weight Dominating Set to detect key events.

(3) DCCI [2]. It is also a graph-based approach that is similar to MWDS but uses the Degree and Clustering Coefficient Information to detect key events.

We use ROUGE and Redundancy as the metrics [5]. The redundancy of an event is defined as the similarity of the two events that have the highest similarity in the key-event set, and the redundancy of the key-event set is the sum of the redundancy of each event.

The results are shown in Table 1, indicating that Post2Event achieves the best performance among all algorithms

Table 1. Performance comparison of key-event extraction

Algorithm	ROUGE-1	ROUGE-L	Redundancy
TF-IDF	0.82	0.47	39.20
MWDS	0.68	0.44	41.48
DCCI	0.68	0.43	44.36
Post2Event	**0.83**	**0.70**	**27.81**

IV. Demonstration

The framework of key event extraction is a web application developed using Python 3.8 and Flask framework. In the demonstration, we will demonstrate the use of some critical functionalities of Post2Event, including Data View and Event View. Figure 2 shows a snapshot of Post2Event.

Data View. Figure 2(a) shows the Data View. It displays the local dataset by default and the popularity of some ongoing datasets crawled by our system. Users can explore the local datasets, check the events, and visualize each data's details in the interface. We show a preview of the ZTE Incident event in Fig. 2(a). Also, Post2Event provides a query interface for users to retrieve relevant events.

Event View. Figure 2(b) shows the Event View. When users select a dataset and click the "View" button, Post2Event will perform the event extraction and display the extracted key events. As shows in Fig. 2(b), Zone 1 shows the key events of "*2020 U.S. President Election*", Zone 2 shows the keywords of each key event, and Zone 3 displays the distribution of the keywords.

References

[1] C. Lin, C. Lin, J. Li, D. Wang, Y. Chen, and T. Li, "Generating event storylines from microblogs," in *CIKM,* , 2012, pp. 175–184.

[2] R. Yuan, Q. Zhou, and W. Zhou, "dTexSL: A dynamic disaster textual storyline generating framework," *World Wide Web*, vol. 22, no. 5, pp. 1913–1933, 2019.

[3] X. Zhao, P. Jin, and L. Yue, "Discovering topic time from web news," *Inf. Process. Manag.*, vol. 51, no. 6, pp. 869–890, Nov. 2015.

[4] Y. Ouyang, B. Guo, M. He, Z. Yu and X. Zhou. Event sensing and vein presentation leveraging microblogging. JOURNAL OF ZHEJIANG UNIVERSITY, vol. 50, no. 6, pp. 1176–1182, 2016.

[5] C. Y. Lin, "ROUGE: A Package for Automatic Evaluation of summaries," Jan. 2004, p. 10.
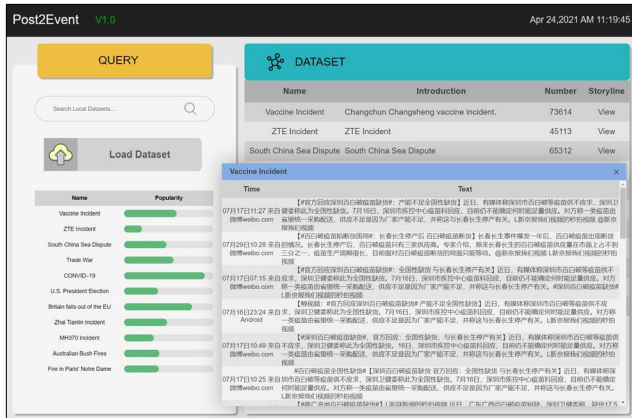
(a) Data View

(b) Event View

Fig. 2. A snapshot of Post2Event