Real-time Identification and Monitoring of Abnormal Events Based on Microblog and Emergency Call Data Using SMART

Jiawei Zhang, Shehzad Afzal, Dallas Breunig, Jing Xia, Jieqiong Zhao, Isaac Sheeley, Joseph Christopher,
David S. Ebert, *Fellow, IEEE*, Chen Guo, Shang Xu, Jun Yu, Qiaoying Wang, Chen Wang, Zhenyu Qian, Yingjie Chen
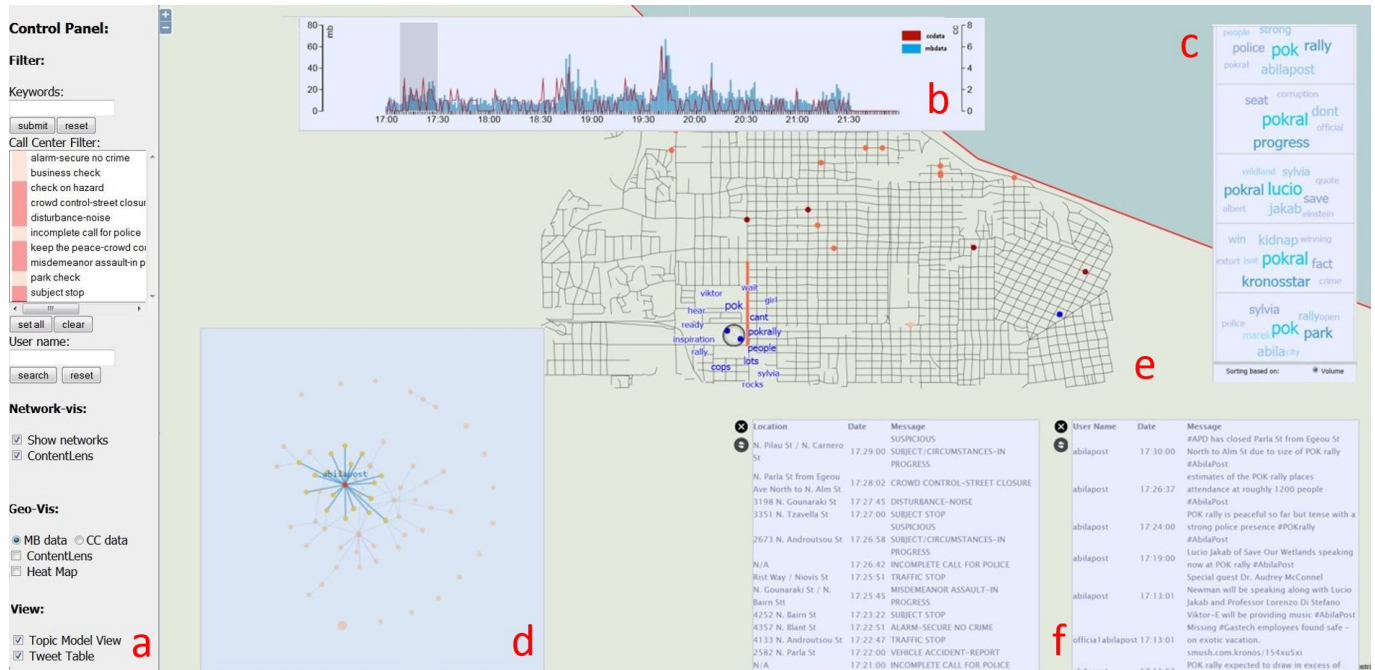
Fig. 1. Our system helps the analyst monitor POK Rally, which is held in Abila City Park. The system consists of a) control panel; b) time series view; c) LDA topic view; d) reply/retweet networks view; e) map view and f) microblog & emergency call table. Through interactive and visual exploration, several major events occurred in the selected time period are revealed, including: 1) Part of Parla St was closed due to concerns of public safety (map view); 2) several key roles appeared in the rally including Sylvia Marek and Lucio Jakab (topic view); 3) some news media, such as *abilapost*, actively posted updates of the rally (networks view).

Abstract— This article describes a real-time visual analytics process based on microblog and emergency call data to solve VAST 2014 Mini Challenge 3. We extended SMART system (Social Media Analytics and Reporting Toolkit), developed by the U.S. Department of Homeland Security's VACCINE Center. Our system consists of multiple linked views to allow the analyst monitor topic evolution, identify influential microblog users, observe geo-location patterns and examine correlations among different data sources. Extensions to our previous work include a time series view, a reply/retweet networks view, and integration of emergency call data.

Index Terms—Visual Analytics, Microblog, Spatiotemporal Analysis, Multiple Linked Views

1 INTRODUCTION

Given real-time microblog and emergency call data stream, it is crucial for investigators to combine multiple data sources to identify abnormal events and build connections among different events in a timely manner. In such scenarios, we summarize the following major tasks which

- *Jiawei Zhang, Shehzad Afzal, Dallas Breunig, Jieqiong Zhao, Isaac Sheeley, Joseph Christopher, David S. Ebert, Guo Chen, Shang Xu, Jun Yu, Qiaoying Wang, Zhenyu Qian and Yingjie Chen are with Purdue University in West Lafayette, IN, USA. E-mail: {zhan1486, safzal, dbreunig, jieqiongzhao, isheeley, christ33, ebertd, guo171, xu537}@purdue.edu, yujun_2005@hotmail.com, {wang1925, qianz, victorchen}@purdue.edu.*
- *Jing Xia is with Zhejiang University in Hangzhou, China. E-mail: xiajing@zjucadcg.cn.*
- *Chen Wang is with Harbin Institute of Technology in Harbin, China.*

serve instrumental roles in the investigation process:

- An integral framework involving analytics of multiple data attributes, including timestamp, geo-location, text (messages) and relationship structure (reply/retweet).

- Effective retrieval of relevant messages from large volume data stream.

- Support of correlation analysis between microblog and emergency call data.

We utilized and extended our previous work, SMART system. SMART is a web-based framework which captures real-time social media stream and allows analysts perform topic extraction, aggregate and visualize major keywords, and categorize messages related to different events. Based on SMART, we incorporated a time series view to visualize data volume trends and a networks view to visualize reply/retweet relationships among microblog users. In the following

sections, we focus on the above mentioned tasks and demonstrate the capabilities of our system to effectively perform investigation analysis.

2 VISUAL ANALYTICS APPROACH

2.1 Multiple Linked Views

In order to understand events of interest to the maximum extent and avoid bias, we utilize the following multiple linked views to visualize a subset of data specified by the analyst.

- **Time Series View** visualizes temporal trends of data volume over time. Microblog data is rendered using blue bar charts and emergency call data is rendered using red line charts.

- **Topic View** visualizes topics extracted using LDA [1] model. Each rectangle represents a topic described by a set of keywords.

- **Networks View** visualizes reply/retweet dynamic networks using force-directed layout [3]. Nodes represent users and edges represent reply/retweet relationships.

- **Map View** visualizes geo-tagged microblog data and emergency call data, the geo-location of which is parsed in the back-end side based on the given KML map file. Microblog records are rendered using blue dots. Emergency call records are parsed as dots or lines based on their semantic structure, and then we manually assign impact factors based on different types of emergency calls and render them using a sequential color scheme, as is shown in the control panel, in Fig.1.

- **Microblog & Emergency Call Table** visualizes detailed microblog and emergency call records including user handle/emergency location, timestamp and message content.

The analyst can easily specify data of interest through interactions including 1) selecting time range in the time series view; 2) clicking a topic in the topic view and 3) selecting one or more users in the networks view using a circular shape. The linked views then automatically refresh to highlight the corresponding data sample.

The analyst can also click a single record in the microblog & emergency call table. The time series view highlights the corresponding time bar. If the record is geo-tagged, the map view highlights the corresponding geo-location using a hot dot animation (Fig.2).
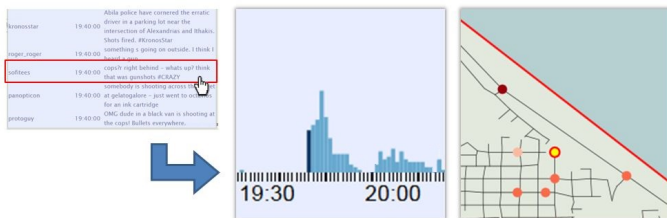

Fig. 2. When the analyst clicks a record in the message table, the time series view and the map view highlight the selected record respectively.

Through multiple linked views, the analyst is able to simultaneously investigate, compare and validate based on different dimensions and perspectives of data stream.

2.2 Locating Relevant Messages in Large Volume Data Stream

Massive data stream with a large portion of irrelevant messages hinders effective identification and monitoring of events of interest. Based on multiple linked views and interactive filtering in individual views, the analyst is allowed to iteratively filter data, refine results and drill down to locate more relevant information.

In the time series view, the analyst can perform filtering by selecting the time range containing volume burst, which can probably indicate a certain abnormal event occurs within that period of time.

In the topic view, the analyst is able to monitor ongoing topics and by clicking a topic of interest, the system triggers filtering based on keywords associated with the selected topic.

In the networks view, assuming users of high degree centrality serve influential roles in disseminating and communicating information in the social network structure, the analyst can perform filtering based on a group of key roles to track their conversation and interaction for more relevant and reliable updates of events.

In the map view, the analyst can also perform spatial filtering to discover location-based patterns.

Hence, effective filtering of individual components and combination of filters supported by the system allow the analyst quickly and accurately capture meaningful information from massive data.

2.3 Support of Correlation Analysis between Microblog and Emergency Call

Microblog and emergency call data have similar attributes, such as timestamp, geo-location and message content (a small portion of microblog messages and all emergency call records are geo-tagged.). We mainly focus on revealing the temporal and spatial correlations through our visual analytics approach.

In the time series view, as is shown in Fig.1, two different types of data have very similar pattern of evolution over time, indicating the occurrence of an abnormal event is usually followed by increasing emergency calls and active posts in microblog traffic. Hence, it is reasonable to regard data volume burst as an early signal for event identification.

In the map view, we render two types of data based on different color schemes (mentioned in Section 2.1). By using contentlens[1], heat map and performing spatial filtering, the analyst is able to restrict data to detailed level of geo-location, combine and examine different data sources to discover location-based correlation.

Incorporating analysis of temporal and spatial correlations, the analyst is able to reveal underlying connections among different events. For example, the shooting & standoff event happened immediately after the hit & run event in the timeline. In both cases the perpetrators were driving a black van and pursued down Egeou Avenue at the same time and in the same direction, which leads to the deduction that the same group of perpetrators engaged in the two events.

3 CONCLUSION

As an extension of SMART system, we present an integrated visual analytics framework to incorporate multiple data sources and data attributes for event identification and monitoring using microblog and emergency call stream. By using the proposed system, we successfully identified several major events happened in Abila city, monitored their timelines, and investigated their underlying connections. Some future work includes more efficient layout and analytics functionality in networks view, incorporating data of multiple social media services for more reliable information retrieval, etc.

REFERENCES

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. In *NIPS*, pages 601–608, 2001.

[2] H. Bosch, D. Thom, F. Heimerl, E. Puttmann, S. Koch, R. Kruger, M. Worner, and T. Ertl. Scatterblogs2: Real-time monitoring of microblog messages through user-guided filtering. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2022–2031, 2013.

[3] T. M. J. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. *Softw. Pract. Exper.*, 21(11):1129–1164, Nov. 1991.

---

[1]Contentlens: A visualization technique applied to SMART, which shows the most prominent terms in geo-tagged messages within a certain area by moving a circular shape. [2]