

Mixed-Initiative Visual Exploration of Social Media Text and Events

IEEE VAST Challenge 2021 MC3 Award for Strong Human-in-the-Loop Analysis Methodology

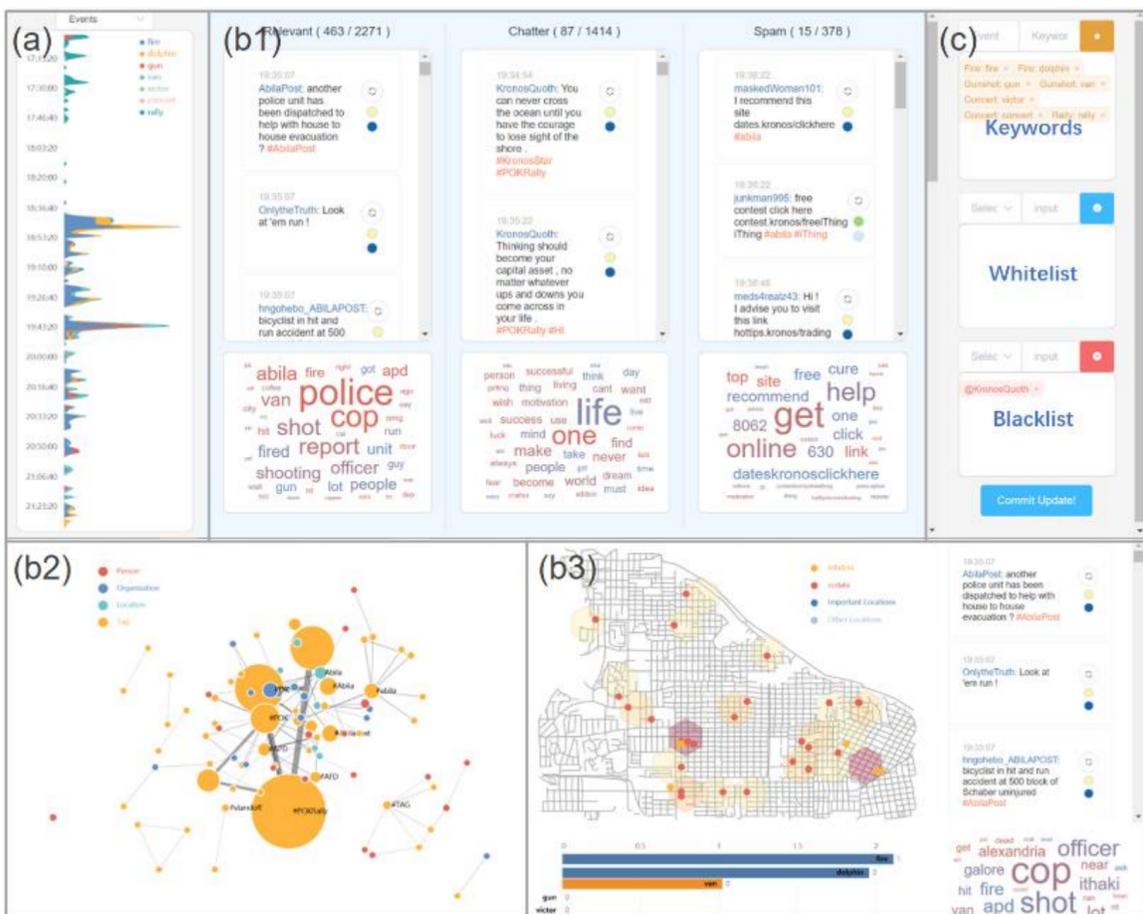
Lei Peng¹ Yuheng Zhao¹ Yijie Hou¹ Qinghong Wang² Siqi Shen¹ Xingui Lai¹
Junting Gao² Jiaqi Dong¹ Ziyue Lin¹ Siming Chen*¹

1) School of Data Science, Fudan University 2) Beijing Data Star Observatory Technology Co., Ltd.

Introduction

IEEE VAST Challenge 2021 Mini-Challenge 3 requires participants to retrospectively detect and evaluate public risks in Abila City during an evening. The provided data are microblog records and emergency dispatch records from local police and fire departments. We design an interactive visualization system which enables interactive classification and mining of social media texts, supports user-defined event analysis, and displays risk of public temporally and spatially. It features mixed-initiative visualization design and human-in-the-loop analysis process. Generally, our visual analytics solution has 2 integrated steps: microblog classification (step 1) and detailed analysis (step 2). The user may first start from an interactive classification and mining process of microblogs to identify relevant message, chatter and spam (step 1). After observing crucial information, he or she may then interactively explore user-defined risk events in the spatial temporal scenarios (step 2). During the whole analysis process, the user's knowledge of the situation deepens through the iterative exploration.

Design



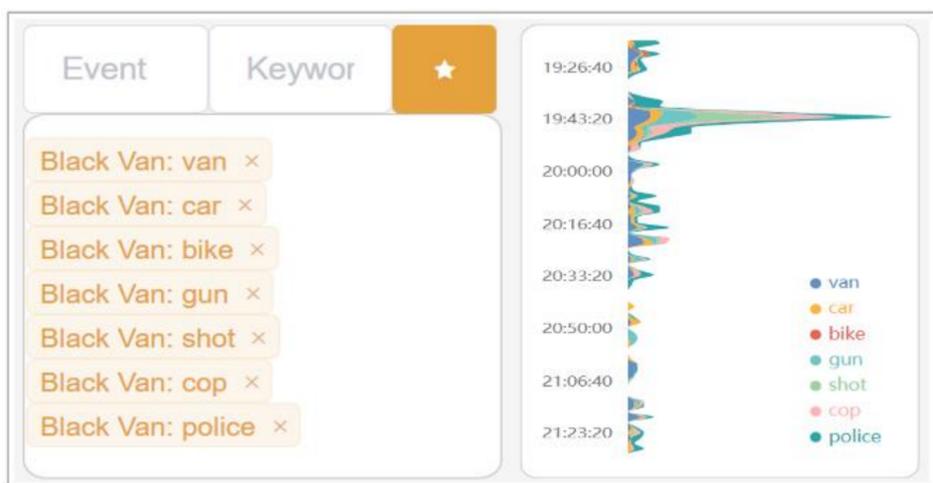
The system interface has a brushable timeline on the left (a), a filter panel on the right (c), and a scrollable central view which can switch between the message view (b1), the entity graph (b2), and the general analysis view (b3).

Timeline (a): All the views are linked to the timeline. And the timeline itself can be rendered based on message type, classification result, user-defined keywords, or sentiments.

Message View (b1): Each category of messages is displayed with its corresponding word cloud. We use a combination of machine learning and human interaction to do the classification.

Black List and White List (c): Users can do batch operations with the help of the black list and the white list. For example, he or she can add bloggers or tags into the black list to prevent them from appearing in the relevant message set.

Entity Graph (b2): From the raw texts, we extract entities such as authors, tags, organizations, locations, and persons to build an entity graph. The entity graph provides event clues and entity relations.



Event Keyword Tool: Having explored with the messages, the word clouds and the entity graph, users may become curious about several keywords of different events that happen during the evening. Users can input self-defined events with corresponding keywords into the event keyword tool. The system will render the timeline with the appearance of keywords, generate a bar chart race under in the general analysis view, and highlight message cards that contain a keyword.



Map: We provide a map of Abila City marked with relevant information. The blue rectangles represent public places that we identify in Abila City. During a certain time period, we mark the locations of microblogs and emergency dispatches on the map. We also define indexes such as risk level and population influence rate in area S during time region T. The higher the estimated risk level is, the redder the area will be coloured.

Acknowledgements

This work is supported by Shanghai Municipal Science and Technology Major Project (No.2018SHZDZX01), ZJ Lab, and Shanghai Center for Brain Science and Brain-Inspired Technology. This work is also supported by Shanghai Science and Technology Commission (Grant No. 21ZR1403300) and Shanghai Sailing Program (No.21YF1402900).

Contact

Siming Chen: simingchen@fudan.edu.cn

