EarthquakeAware: Visual Analytics for Understanding Human Impacts of Earthquakes from Social Media Data

Shuai Chen[1*] Sihang Li[1] Liwenhan Xie[1] Yi Zhong[1] Yun Han[1] Xiaoru Yuan[1,2†]

1) Key Laboratory of Machine Perception (Ministry of Education), School of EECS, Peking University
2) National Engineering Laboratory for Big Data Analysis and Application, Peking University

Abstract

In this paper, we propose EathquakeAware, a visual analytics system for social media visualization. Our system supports analyzing geo-tagged messages from different perspectives, including location, keyword, account. It can summarize the evolution patterns of topics and reveal the discussing trend of different keywords to help evaluate situations in different regions. We take VAST Challenge 2019 Mini Challenge 3 as case studies to demonstrate its effectiveness in a fictitious analysis scenario.

Index Terms: Human-centered computing—Visualization—Visualization application domains—Visual analytics

1 Introduction

Social media has been an important platform for users to report events in their life. The geo-tagged messages reported by users can increase situational awareness of the reported areas and help analysts make decisions to deal with events [1]. However, there is often much noise in public messages, and it is not easy to extract critical information and analyze them in the large volume of messages.

In this paper, we propose EarthquakeAware, a visual analytics system to support data investigations from different aspects, including location, time, keyword and people. Our system first provides an overview of the evolution of topics in the messages. Then users can check the spatio-temporal trend of a keyword. Users can further investigate the messages containing the keyword in detail to evaluate the situations.

2 Overview

Throughout this work, we use the synthetic dataset from IEEE VAST Challenge 2019 Mini Challenge 3. It is a selected record of posts on a social media platform in a city when earthquakes happen. The once-vibrant city, St. Himark, suffers from cracked roads, nuclear leakage from the power plant and so on from the earthquake. To guide a better emergency service, it is of important concern for the city officials to have an up-to-date view for the city officials of the structural and humanitarian impact caused by the earthquake on a neighborhood-by-neighborhood basis.

There are 41,941 records in total with four attributes, i.e., account, location, time, and message. Specifically, our tasks are as follows. 1) Characterize conditions across the city and recommend allocation of resources 5 hours and 30 hours after the earthquake. 2) Identify times when conditions change in a way that warrants a re-allocation of city resources. 3) Understand how has the earthquake affected life in the city.

To support the above tasks, we choose to focus on the analysis of the spatial-temporal occurrences of the messages. We identify the following design goals to be achieved in our system.

*e-mail: shuai.chen@pku.edu.cn
†e-mail: xiaoru.yuan@pku.edu.cn

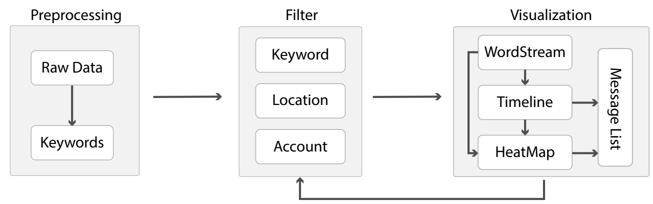

Figure 1: Pipeline of EarthquakeAware.

A. Overview of topics evolution. For characterizing conditions across the city, a visualization should tell users what people discuss during the earthquakes and how their discussions change with time.

B. Events identification. When analyzing messages posted by citizens, it is crucial to identify key events from their messages and where these events happen so that appropriate measures can be taken to re-allocate the city resources.

C. Spatial-temporal analysis. Different locations are affected by earthquakes with different levels. It is important to find the locations which need resources most. When conditions change, a visualization should help reveal the change.

D. Visual clutter reduction. It is challenging to reduce visual clutter when facing a huge volume of data. EarthquakeAware aims to provide a clear view of events in the messages and flexible interactions to filter messages.

3 Visual Interface

The overall design is based on WordStream [2]. Five views are designed to finish the tasks according to the design goals.

Entity View shows the frequency of different locations, keywords and accounts in the messages (Figure 2a). It highlights active locations, accounts and hot keywords people discuss (design rationale **B**). By clicking a bar, users can filter the messages to analyze messages only containing this entity. For example, users can click a bar in the location barchart, and all other views will be updated to help investigate the situation in this location.

WordStream visualizes the evolution of keywords in the messages with time (Figure 2b). Stop words are removed to reduce noise in the data. In each slot, the number of keywords shown is determined by the volume of messages. The size of the keywords encodes the frequency. With this view, users can quickly find new events with the burst of the specific keywords (design rationale **A, B**). By clicking one keyword, it will be highlighted in different time slots. The Timeline and Heatmap will also be updated. A slider is provided to help users change the time slot size. There is a trade-off between overview and details for different time slots sizes.

Timeline shows the temporal distribution of keywords (Figure 2c). Users can check different keywords by clicking them on the WordStream or manually typing in. The aligned keywords timeline help users easily compare the temporal distribution of keywords (design rationale **C**). Users can brush the timeline to investigate the spatial distribution of keywords on HeatMap in the selected time range.

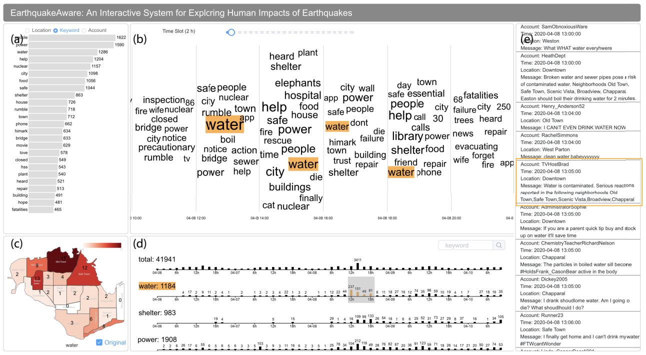

Figure 2: The interface of EarthquakeAware. (a) Entity View; (b) WordStream; (c) Timeline; (d) Heatmap; (e) Message List.

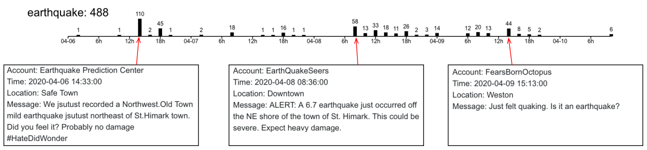

Figure 3: Messages about earthquakes.

HeatMap visualizes the spatial distribution of the filtered messages and provides a spatial context for users (Figure 2d). The heatmap help users easily compare activeness of different locations (design rationale **C**).

Message List shows the original messages, including account, location, time and message content (Figure 2e).

As shown in Figure 1, EarthquakeAware's system consists of three components. First, in the preprocessing step, keywords were extracted from the messages. Then, these messages could be filtered by location, keyword or account. The filtered messages would be rendered in the visualization component. The views in the system are linked to help users identify events and check details (design rationale **(**D)).

4 Case Study

We use two cases to illustrate the effectiveness of EarthquakeAware. The first one is about the discussion of the happening of the earthquakes, and the second one is about the damages caused by the earthquakes.

4.1 Happening of Earthquakes

By checking the occurrence of "earthquake" in the messages, we found three peaks in the timeline (Figure 3). For each bar in the timeline, messages contained the keywords were further analyzed. At 14:33:00, on April 6, 2020, Earthquake Prediction Center posted that it recorded an earthquake. However, many people reported that they did not feel the earthquake. So, we thought it was the first earthquake, which was not severe. At 08:36, two days later, EarthQuakeSeers posted the alert that a heavy earthquake hit the city, which might cause heavy damages. It is the second earthquake. At 15:13:00, on April 9, people reported that they felt shaking, which meant the third earthquake happened. As the earthquake happened in the north-eastern part of the city, Safe Town reported messages about "earthquake" most (Figure 4a).

4.2 Damages Caused by Earthquakes

The earthquakes have caused significant damages to the citizen's life. The WordStream provides users a good overview of the keywords discussed at different time. We found that on April 8, there was a burst of messages on social media and abnormal keywords

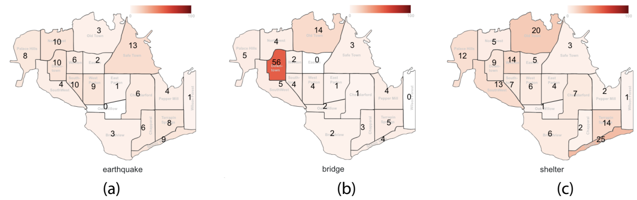

Figure 4: Spatial distribution of keywords "earthquake" (a), "bridge" (b), "shelter" (c).

like "water", "power", "help" and "shelter" appeared (Figure 2b). By clicking these keywords, the Timeline view updated and it proved that people discussed these words more often than before (Figure 2d). The Message List shows detailed messages about water (Figure 2 e). For example, some people reported that water was contaminated. The abnormal distribution of keywords helps us quickly find what happened to the citizen's life.

The HeatMap helps users evaluate the emergency of events in different places. From the HeatMap, people in Down-town posted the most messages about closeness and opening of bridges (Figure 4b), but bridges were only at the boundary of the city. An account was very active in reporting the condition of bridges, which might be responsible for bridge management. Although most people live in Down-town, Scenic Vista reported most messages about "shelter" (Figure 4c). By checking the Message List, we found that many people went to find shelters as the earthquakes destroyed their original living places. There were schools and libraries opened as shelters in Old-Town and Scenic Vista.

5 Discussion

Our current system only supports the analysis of static data. However, in a real emergency, we are faced with streaming data. To improve the usability of our system, the WordStream can be adapted to visualize stream data, which requires the system to extract keywords in the stream effectively. There are also uncertainties in the data. In HeatMap, we only show the density of the original messages. As some messages reposted other messages with different locations, we think they could not reveal the real situations in these areas.

6 Conclusion

In this paper, we present a visual analytics system for microblog messages investigation. Our research emphasizes a multi-filter visual design. In the future, we would try to combine advanced text mining methods to extract topics from the messages effectively instead of keywords. Besides, we would extend the current calculation and visualization methods to support stream data analysis.

Acknowledgments

The authors wish to thank the IEEE VAST Challenge committee and the anonymous reviewers. This work is funded by the National Key Research and Development Program of China (2016QY02D0304) and the National Program on Key Basic Research Project (973 Program) No. 2015CB352503. This work is also supported by PKU-Qihoo Joint Data Visual Analytics Research Center.

References

[1] S. Chen, L. Lin, and X. Yuan. Social Media Visual Analytics. *Comput. Graph. Forum*, 36(3):563–587, 2017. doi: 10.1111/cgf.13211

[2] T. Dang, H. N. Nguyen, and V. Pham. WordStream: Interactive Visualization for Topic Evolution. In *EuroVis 2019 - Short Papers*, pp. 103–107, 2019. doi: 10.2312/evs.20191178