

Safeguarding Abila: Real-time Streaming Analysis

VAST 2014 Mini Challenge 3: Streaming Text Analysis

Patrick Butler*

Parang Saraf†

Naren Ramakrishnan‡

Discovery Analytics Center
Department of Computer Science
Virginia Tech

ABSTRACT

We introduce a system for visualization and analysis of geo-spatial and temporal data from call center and microblog sources. We provide a streaming client for interacting with the data in real time. We present the data in a partitioned format using coordinated visualizations that allows the analyst to view the data in multiple dimension simultaneously. This allows the user to see patterns that occur in space and time. This project was developed in response to the VAST 2014 Mini-Challenge 3.

Index Terms: H.5.2 [Information Interfaces and Presentation (e.g., HCI)]; User Interfaces—Interaction styles (e.g., commands, menus, forms, direct manipulation)

1 INTRODUCTION AND PROBLEM OVERVIEW

The VAST 2014 Mini Challenge 3 describes a hypothetical scenario where some of the employees of an imaginary organization, GASTech have gone missing and it is speculated that an environmental activist group, Protectors of Kronos (POK), is responsible behind the disappearance. The provided dataset includes several hours of microblog data as well as call center data which comes from the public services departments. The challenge requires identification of suspicious activities and behavior hidden in data. Additionally, it is also required to prioritize all such activities in order to determine suspicious persons and locations that are worthy of reporting to law enforcement agencies.

We developed a web-based visual analytics system to analyze geospatial, temporal and microblog data. The system, which was developed from scratch over the course of the competition, empowers an analyst to uncover any aberration inherent in data. It provides several widgets to analyze geotagged microblog data, retweet frequency and term frequency.

2 SYSTEM DESIGN

A dynamic website was developed using websockets for interfacing with a backend and `d3.js` [1] and `nvd3.js` for the visualization frontend. The geospatial visualization was created using the `gmaps.js` interface to a Google Maps-based visualization.

For debugging and testing purposes a server was built which could cache and store streams played by the contest provider's server. This server was built in Python and allowed the team to play streams at 100 times the original rate which made rapid prototyping and debugging much easier. This replay system was not used during the analysis of the Stream 3.

*e-mail: pabutler@vt.edu

†e-mail: parang@cs.vt.edu

‡e-mail: naren@cs.vt.edu

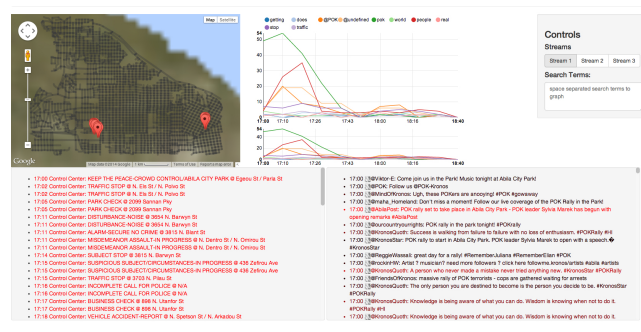


Figure 1: The interface for our Mini-Challenge 3 solution. The interface is split into four quadrants: geospatial view, term frequency graph, call center messages, and microblog messages.

The visualization is split up into four panels as seen in Figure 1: a geospatial visualization, a term frequency graph, the call center message panel, and the microblog message panel.

The geospatial visualization provides a map that allows for the visualization of individual microblog entries that are geotagged. This enables the analyst to easily see patterns in space. The text of individual posts can be seen by hovering mouse over the markers representing the entries.

The term frequency graph displays individual terms. By default these terms are the most common terms that appear in the messages. The analyst can instead choose a list of arbitrary terms to graph in order to further explore the data.

The call center message panel displays the call center messages filtered by the terms selected by the analyst.

The microblog message panel is similar to the call center message panel, however it offers several extra options. As with the call center message panel microblogs can be filtered by the analyst; however, the microblog entry is also color coded such that the redder a message is the more times it has been reblogged. Since reblogs are not a part of the message format reblogs are detected using string expression matching. Microblog messages sometimes incorporate geotagged information which can be used to highlight the pertinent marker in the geospatial visualization. If geotagged information is available the icon is displayed in color and may be clicked to activate the animation; otherwise it is displayed in gray.

3 ANALYTICAL PROCESS

For Mini Challenge 3 we were asked to find several events of interest. In each case the toolkit we developed helped immensely in determining events. For each case the main action of discovery was typing in a few relevant keywords and looking at their occurrence in both the microblog and the call center messages.

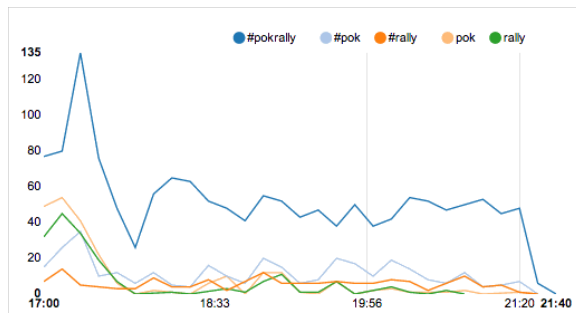


Figure 2: Term frequency of related words to the POK rally.

3.1 POK Rally

Our first event was found by searching for the term “POK”. This term was chosen as it was an entity of interest in the previous mini challenges and in the problem statement. This search immediately brought up tweets including information about the POK rally. Through the visualization we can see the location of the protest as well as follow the activity occurring in relation to the protest via the microblogging service. Figure 2 shows how mentions of the POK rally pick up around the beginning of the rally and continue to be a topic of discussion even after the rally ends.

3.2 Rumor Spreader and Troll

Our second event of interest was found by searching for the term “GasTech”. This brought up messages from a user @OfficialAbilaPost which said: “Missing # Gastech employees found safe - on exotic vacation. smush.com.kronos/154xu5xi”. Due to the color coding of this message we could immediately tell it was reblogged only twice, a small number of times for such momentous news. Further examination of the microblog messages produced by this user show contradictions to microblog entries by other users including: @AbilaPost and @KronosStar who are reporting and getting reblogged, and thus likely to be legitimate. Furthermore, @OfficialAbilaPost makes the claim that “Elian discovered alive, living in Tethys” which is verifiably untrue. We thus decided to discount this user as a troll.

3.3 Fire at the Dolphin Apartments

While researching the troll we found mentions of a fire in downtown Abila. This troll while spreading misinformation actually gave us pointers to actual events. The troll while always posting false information, often posted information relating to real and verifiable events. One such example was “#AFD breaking scandal: setting fires in rundown neighborhoods for training exercises, accidental urban renewal. smad.co.kronos/383xaw”. From this post we decided to look deeper into the fire and searched simply for “fire” and “afd”. The following is a timeline from the events around the fire.

- 18:40 : A fire was reported near N. Achilles St / N. Madeg St.
- 18:42 : Abila Fire Department (AFD) dispatches 2 units and also ask the police department to send units for crowd control
- 18:45 : Fire trucks arrive at the building shortly followed by the police
- 18:47 : Request for medical staff along with ambulance and an additional fire truck is placed
- 18:52 : Firefighters assist residents out of the building. The building has six floor and floor 2 caught fire
- 18:53 : AFD orders evacuation of nearby buildings
- 18:59 : Firefighters report rescuing a resident trapped on upper floor
- 19:13 : More fire trucks followed by additional police arrive
- 19:22 : AFD announces expansion of evacuation area
- 19:35 : Police starts doing house to house evacuation
- 20:05 : A firefighter is injured and is pulled out of flames and is taken to Abila Hospital by paramedics

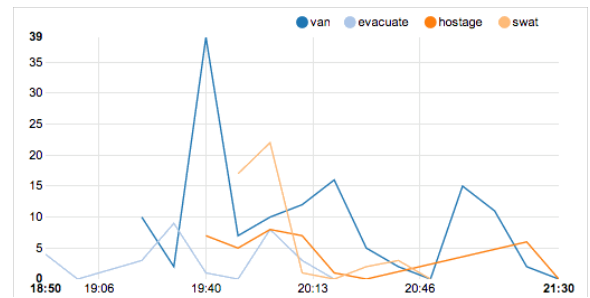


Figure 3: The term frequency graph of search terms related to the black van and the following shootout.

- 20:25 : The fire appears to be under control
- 20:40 : The fire suddenly seems to be escalating, more units called
- 20:50 : The fireman turn their hoses onto the surrounding buildings, in order to prevent fire from spreading
- 21:00 : Top floor of the building collapsed
- 21:30 : There is an explosion at the Dancing Dolphin Fire

3.4 Suspicious van and the resulting hostage standoff

The next event of interest was discovered after reading about a shootout at the Gelato Galore while skimming through the tweets. The shootout involved a possible hostage, and a black van so we searched: “van evacuate hostage swat”. This aided us in tracking down more information about the van and that it was likely this van had been involved in a hit and run earlier in the evening. A summary of this mystery van is as follows.

- 19:20 : A message reported that a black van hit a persons car near N Souliou St 1424 1498
- 19:22 : Another message reported that a black van hit a guy on bike near N Edessis St / N Menelaou, which gets reported and police starts chasing the van
- 19:40 : After a brief chase, the van is cornered-off in the parking lot near Gelato-Galore. Both the cops and the van occupants open their doors and started firing at each other during which an office is shot
- 19:42 : Request for ambulance and additional police units is sent
- 19:55 : SWAT arrives at the location
- 20:00 : The guy in the van says that he will shoot the hostages if they dont let them go
- 20:06 : Meanwhile, police starts evacuating near by businesses
- 20:10 : A police negotiator starts talking to the guy
- 21:00 : The guys goes back in the van and have an animated argument with the other van occupant.
- 21:17 : They decide to surrender and come out with hands up
- 21:18 : Police put them (a male and a female) in a waiting car and rescue the hostages (2 females)

Additionally, the term frequency graph for this search appears in Figure 3. From this graph we can see that the hostages aren’t connected with the van until later in the day.

4 CONCLUSION

VAST 2014 Mini Challenge 3 presented a problem that requires analysis of semi-structured textual data from microblogs and a fictional call center, adeptly analyzed using the visual analytics tools presented here.

ACKNOWLEDGEMENTS

We would like to thank Ritika Dokania for her creative inputs and feedback on the visualization, as well as for lending her voice to the explanatory video that describes the system. This work is partially supported by US NSF grant CCF-0937133.

REFERENCES

- [1] M. Bostock, V. Ogievetsky, and J. Heer. D: Data-Driven Documents. *IEEE Trans. Vis. Comput. Graph.*, 17(12):2301–9, Dec. 2011.