City Research Online

City St George's, University of London

Citation: Xu, K., Nguyen, P. & Fields, B. (2015). Visual Analysis of Streaming Data with SAVI and SenseMAP. 2014 IEEE Conference on Visual Analytics Science and Technology (VAST), ISSN 2325-9442

This is the draft version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/19059/

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

Visual Analysis of Streaming Data with SAVI and SenseMAP

Kai Xu, Phong H. Nguyen, and Bob Fields

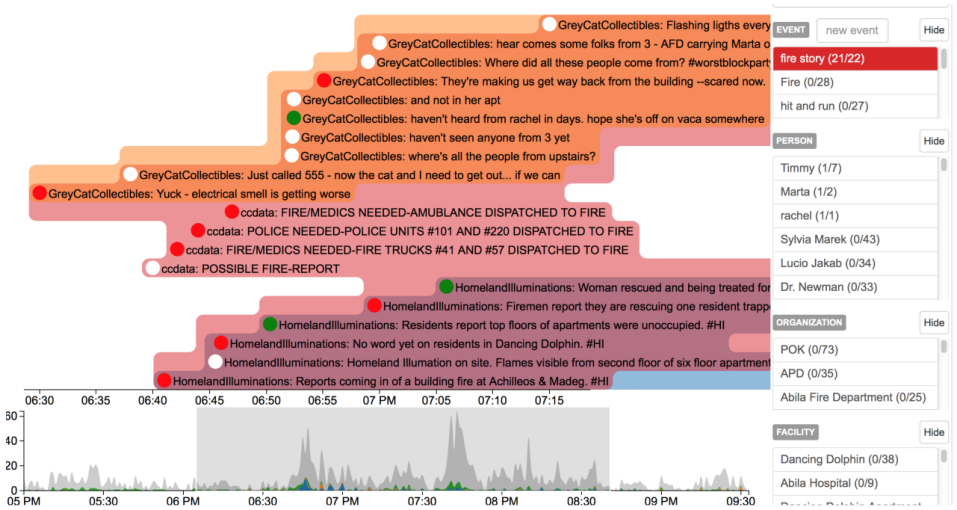

Figure 1: The timeline view in the Social Analytics VIsualisation (SAVI). The messages (top-left part) are grouped according to their topics, indicating by background color. The histogram at the bottom and the entity list on the right can be used to filter the messages displayed by time interval and entity instance such as person and organization respectively.

1 Introduction

Two tools were developed for the analysis tasks in the VAST Challenge 2014 Mini-Challenge 3: Social Analytics VIsualiszation (SAVI) and Sense Making with Analytic Provenance (SenseMAP).

The SAVI provides linked timeline and geospatial view of the message streams. It allows search and filter based on text analysis, such as named entity extraction and sentiment analysis, and enable analyst to identity interesting events, tracking them temporally and geospatially, and find related information.

The SenseMAP allows analysts to record finding, together with the visualisations that lead to it, and generate hypotheses and construct narratives. The entire sense making process is captured so it is possible to go back to any previous reasoning state and share or playback the process.

The fact that complex investigations are often carried out by several people working collaboratively is supported in our tools by allowing several people to use SAVI concurrently on the same dataset. Used in this mode, each users analyses are kept separate. However, the SenseMap tool supports collaboration and sharing of hypothesis and narrative structures. Each users findings are available to others, and the findings of different users can be compared, combined, merged, and so on, to produce a collective analytic result.

The key insights that informed the design were firstly, that it was important to see and interact with dataset from different perspectives, and secondly that an essential part of an investigation is the construction and representation of a narrative.

A video showing both tools are available online: `http://invisque.com/timesets/mc3.mp4`.

2 SAVI: Social Analytics VIsualisation

SAVI is a web-based multi-user tool that allows near real-time exploration of the MC3 data stream. It has two coordinated views: timeline and geospatial.

Fig.1 shows the timeline view in SVAI. At the bottom is a histogram showing the frequency of the messages over the time. On the right-hand side is a collection of "entites" identified from the entire collection of messages. The entities are grouped based on their types such as "person" and "organization". Both the histogram and the entity list also act as filters. User can selective an time interval on the histogram, and as a result only messages within that interval will be shown in the timeline. Similarly, users can select one or more entities and only message containing such entities will be shown on the timeline. Each selected entity will be shown with a distinguish color, and the messages containing the entity in the timeline will have matching color background. For example, in the configuration shown, the view is displaying messages filtered by three different entities. These groups of messages are highlighted with different colours, and it is possible to see the overlaps between the different groups. On the timeline, messages are arranged in time order along the x-axis. The dot at the beginning of each message shows its time. The dot color indicates message sentiment: red for negative, green for positive, and white for neutral. Messages are grouped according to the entities they contain, if there is more than one entity selected. Groups with overlapping messages, i.e., messages containing more than one entities, are placed next to each wherever possible.

Some of the messages have geographical locations, given as lat/long positions, and others have street addresses that can be turned into approximate positions using the Abila street map data. These can be shown in the SAVI map view, as shown in Fig.2. Messages appear as dots on the map, colour coded to indicate the time they where sent (a key in the upper right of the view shows the time range from the earliest to the latest of the messages displayed, and the range of colours from red oldest to green newest). Dots in the map can be selected (highlighted in blue), and the lower portion of

the map view shows the selected messages with additional details. This map view is linked to the timeline view, so only those messages shown in the timeline and have geolocation are displayed on the map.

Figure 2: The geo-spatial view in SAVI. Messages are shown as dots and colored according to time (legend in the top-right corner). The details of the selected messages are shown as a list at the bottom.

From our experience, identification of interesting events can start from several places. Peaks in the histogram shows a burst of messages, usually the result of some significant events. We can often discover relevant events by focusing on such short period. Similar analysis is quite effective for messages with geo-locations. By studying the area with large number of geo-tagged messages, we were able to identify the fire at the Dancing Dolphin apartment and the shooting at the Gelato Galore shop. Such analysis is less successful for popular entities, such as persons or organizations, as the correlation between popular entity and interesting event is not strong. We also found in some cases it is necessary to expand search to find more relevant information in order to identify interesting events. In the case of "black van", after it was mentioned in a microblog message, we searched for all messages containing blank van and uncovered its erratic driving route.

3 SenseMAP: Sense Making with Analytic Provenance

The SenseMap is designed to support sensemaking, more specifically the organisation of findings and the generation of hypothesis and narrative. Fig.3 shows the interface of the SenseMAP. On the left is a list of findings, which can be added to the "reasoning space" on the right for further analysis. Within the reasoning space, each finding is shown as a box containing the screenshot that led to it (top) and its description (bottom). The findings can be further linked to form a narrative describing a complex scenario. In case of hypothesis, findings can be linked as evidence supporting or contradicting the argument. At the bottom is the provenance of the reasoning process. Each nodes is a reasoning space states and they are joined by "actions" that change the state to the next one. Actions include not only adding/removing findings/edges, but also user interaction such as repositioning the findings. As a result, it is possible to replay the reasoning process, go back to any previous reasoning state and branching from these, and share the process with other users.

4 Implementation

Both SAVI and SenseMAP are web applications. Node.js (`http://nodejs.org/`) is used as the back end server with MongoDB (`http://www.mongodb.org/`) as the data storage, which is updated as new data come in from the stream. Socket.io (`http://socket.io/`) is used for the client-server communication and the front end visualisation is mostly done with D3 (`http://d3js.org/`). The text analysis is achieved through web service provided by AlchemyAPI (`http://www.alchemyapi.com/`).

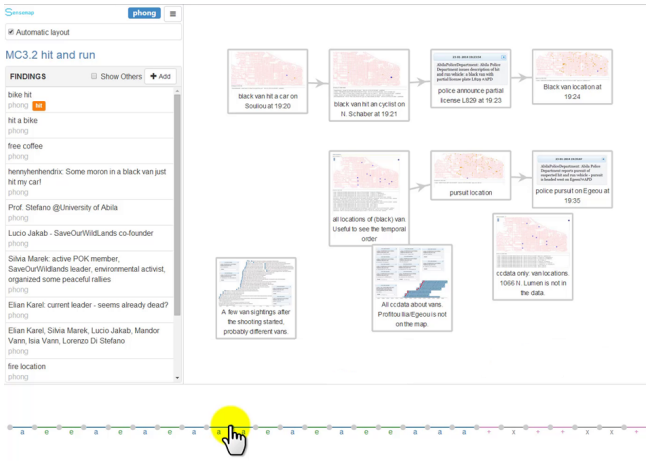

Figure 3: SenseMAP: on the left is a list of "findings" user created. These can be used to construct narratives in the workspace on the right.

5 Reflection

Lessons learned include the following:

- Having multiple coordinated views proved very valuable.
- Although the size of the dataset was relatively modest in this challenge, the mechanisms that SAVI provided for managing the scale of the data were invaluable. Especially useful was the combination of named entity extraction and the filtering tools provided on the timeline view.
- Although only a fairly small proportion of microblog messages came with location data, these messages, combined with the emergency call logs about some of the most significant events of the scenario, made geolocation and the map view very useful. The ability to see events mapped in space (as well as in time, in the timeline view), and the coordination of filtering across these views provided an excellent way of gaining insights into the scenario.
- Furthermore, the extent to which colour coding of blobs on the map to indicate time, and the density of colour to indicate the density of tweets allowed the identification of significant points in space and time.
- Similarly, the way the histogram gave an overview of microblog activity, and more detail in the timeline gave the analyst multiple levels of details of the information.
- The SenseMap made it possible to collect and organise findings, reducing user's cognitive load. The visual representation of the findings in the reasoning space made it easy to see the links between findings and facilitated the comparison of alternative hypotheses and narratives.

While SAVI provides good search and filtering, one thing still missing is the ability to finding related information without affecting the current analysis. For example, it will be nice to be able to quickly find information about "blank van" when investigating the shooting at Gelato Galore, without having to search a new line of

analysis and come back to the shooting later. While the reasoning space in the SenseMAP provides completely freedom in term of positioning and linking findings, the next step is to introduce new features specifically for argumentation (for example with predefined argumentation template) and narrative construction (for example automatic layout for narratives).