

Visual Analytics with Storyboarding to engender multivocality and comprehension of Microblog data for Crisis Management

Extended Abstract

NATO/PfP UNCLASSIFIED

*Dr. Jonathan C. Roberts (principle author)*¹ <j.c.roberts@bangor.ac.uk>

*Dr. Llyr ap Cenydd*¹ <llyr.ap.cenydd@bangor.ac.uk>

*Dr. Panagiotis Ritsos*¹ <p.ritsos@bangor.ac.uk>

*Mr. Richard George*¹ <r.l.george@bangor.ac.uk>

*Dr. William Teahan*¹ <w.j.teahan@bangor.ac.uk>

*Dr. Rick Walker*² <rick.t.walker@gmail.com>

¹ Visualization & Medical Graphics Group, School of Computer Science, Bangor University, UK

² Department of Computing and multimedia technology, Middlesex University, UK

Relevance to symposium.

We have developed a visual analytic interface that visualizes microblog data, and specific views can be refined such to create a visual storyboard of multivocal trends. Our motivation is to enable an explorative interface for crisis management.

Symposium topics (listed in CFP): information trails, analytical reasoning.

Content

There are many opportunities for users to discover trending events and unfolding scenarios, through the evaluation and visualization of large social network data. Indeed, social networking has become an intrinsic part of modern society, and therefore by harnessing the information within Twitter it is possible to understand particular moods of certain demographics, alerting authorities to events before they happen and ultimately, preventing events such as the London riots of 2011.

The small size of the microblogs leads to a high frequency of posts, and a tendency to post during or immediately after an event rather than several days later [1] leads to the dynamic nature and timely importance of this data. Disaster response agencies wish to take advantage of this short latent period between event and post to help inform and direct their efforts in the event of a crisis [2], and although rumors and misinformation is present in microblogging it is possible to estimate their reliability [3].

However, making sense of this information is still a challenging task, especially during pressure-filled situations [4]. The area of microblogging, in particular, provides large amounts of data on a daily basis. When the task is to understand how a crisis is unfolding, the important information is buried within much irrelevant, and often contradictory data.

The creation of corpuses can help: but often they can take hours to create and often rely on typical or well-formed language, rather than the informal language that is found on the microblogging sites. Furthermore, there may be different interpretations over how this information is unfolding: the analysis of microblog data is often uncertain, and analysts need to create different potential pathways conclusions. The use of part-of-speech analysis can also help [5] tease out the interesting stories, and focus on more meaningful data. But there are still many potential outcomes to investigate and different stories (or multivocal ideas) to report. Consequently, we believe it is important to both visualize this information and also present it in a way that engenders discussion and reports these ‘many-voices’.

In our work, we have created a visual analytic tool (epSpread) that visualizes microblog data [6]. The user can analyze different facets of the information. We utilize an in-data method to create an appropriate corpus, and then perform language processing on the microblogs. Through this strategy we are able to analyze the data in a timely manner (rather waiting for hours until the NLP algorithms complete). We also use a storyboarding design that enables visual depictions of the microblog data. The full visual interface contains several different views that are linked together – this enables the analyst to drill down to the specific information. Then, when the analyst has a suitable visualization of the data at a particular time-point, they can pin the information to the timeline. This creates a thumbnail of their visualization (actually what is created is a zoomed view of one of the main panels). The user can then explore another situation, and pin that one, etc. A user can ‘pin’ several different visual depictions to the timeline. In fact, these individual small views are fully functional visualizations, and the user can still manipulate the view in this state (albeit with very fine mouse pointer manipulations). Because it is possible to pin several visualizations we utilize a powerwall setup. This is a tiled display of several monitors. In addition the user can then zoom into one particular view, and then zoom out to see all the pinned views.

- [1] JAVA A., SONG X., FININ T., TSENG B.: Why we twitter: understanding microblogging usage and communities. In *Proc 9th WebKDD and 1st SNA-KDD 2007* (New York, NY, USA, 2007), ACM, pp. 56–65.
- [2] KIREYEV K., PALEN L., ANDERSON K.: Applications of topics models to analysis of disaster-related twitter data. In *NIPS Workshop Appl. for Topic Models: Text and Beyond* (2009).
- [3] CASTILLO C., MENDOZA M., POBLETE B.: Information credibility on twitter. In *Proceedings of the 20th inter- national conference on World wide web* (NY, USA, 2011), ACM, pp. 675–684
- [4] THOMAS J. J., COOK K. A.: A visual analytics agenda. *IEEE Comput. Graph. Appl.* 26 (January 2006), 10–13.
- [5] PRITCHARD I.C., WALKER R., ROBERTS J.C., Visual Analytics of Microblog data for pandemic and crisis analysis. In *Int. Workshop on VA, EuroVA 2012*.
- [6] AP CENYDD L., WALKER R., POP S., MILES H., HUGHES C., TEAHAN W. J., ROBERTS J. C.: epspread - story- boarding for visual analytics. In *IEEE Conference on Visual Analytics Science and Technology, Providence, Rhode Island, USA* (2011), pp. 311–312.