International Workshop on Visual Analytics (2012)
K. Matkovic and G. Santucci (Editors)

Exploring Patterns of Uncertainty in Crowdsourced Crisis Information

Iain Dillingham, Jason Dykes and Jo Wood[1]

[1]giCentre, School of Informatics
City University London

The humanitarian community is reluctant to use reports from social media when responding to a crisis event, as it fears the costs of untrustworthy and inaccurate information [TBJ*11]. Organisations such as Ushahidi use crowdsourcing to identify 'reliable' reports, but this introduces further uncertainty. By combining human and computational approaches, VA offers a powerful means of tackling complex socio-technical issues such as these. We present an ongoing research programme that combines interactive visual representations with spatial statistical functions to explore patterns of uncertainty in crowdsourced crisis information.

Our prototype software allows us to explore the spatial, temporal and thematic distribution of crisis event reports through linked views that incorporate spatial statistical functions. For example, we can construct a standard ellipse to summarise the spatial distribution of reports from a given time period (Figure 1). We can then step through time periods to see how this distribution changes. We plan to incorporate multiple selections to allow for comparisons between different areas at the same time periods, or the same area at different time periods.

Crisis event reports have a 'locality description' that volunteers geocode to produce geographic coordinates. In previous work, we found that most locality descriptions refer to features, such as named places, rather than to distances and directions from features [DDW12]. Potentially, uncertainty varies between types of locality descriptions, so we could, for example, expect features to be located more accurately than features combined with distances or directions. By incorporating text mining techniques, we hope to test this hypothesis and investigate the relationship between locality descriptions and geographic coordinates, and explore how this relationship varies over space and time.

We have applied our software to a dataset relating to the 2010 earthquake in Haiti and note the even spatial distribution of reports for the duration of the dataset (Figure 1). This is unexpected because the underlying characteristics, such as population and topography, are clearly non-uniform. However, this even spatial distribution hides variation at finer temporal resolutions: For example, each of the first seven days of the dataset are characterised by greater dispersion in an approximate NE–SW direction. We plan to extend our analysis to related datasets and are developing a methodology for iterative, analysis-guided VA software development.

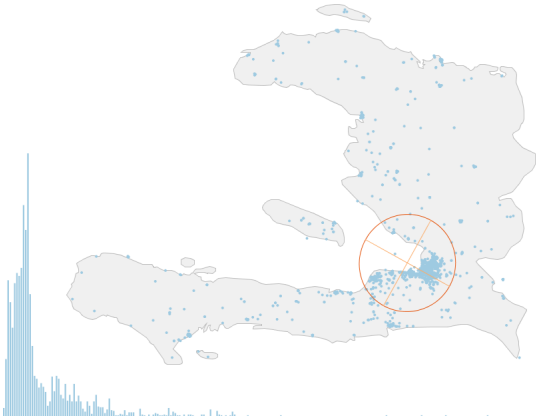

Figure 1: *Summary of the spatial distribution of reports for the duration of the Haiti dataset.*

References

[DDW12] DILLINGHAM I., DYKES J., WOOD J.: Characterising locality descriptions in crowdsourced crisis information. In *Proceedings of the GISRUK 20th Annual Conference* (2012). 1

[TBJ*11] TAPIA A. H., BAJPAI K., JANSEN J., YEN J., GILES L.: Seeking the trustworthy tweet: Can microblogged data fit the information needs of disaster response and humanitarian relief organizations. In *Proceedings of the 8th International Conference on ISCRAM* (2011). 1