Characterising Locality Descriptions in Crowdsourced Crisis Information

Iain Dillingham, Jason Dykes and Jo Wood

giCentre, School of Informatics, City University London, EC1V 0HB
Tel. +44 (0)20 7040 295, Fax +44 (0)20 7040 8584
iain.dillingham.1@city.ac.uk, j.dykes@city.ac.uk, j.d.wood@city.ac.uk

Summary: Humanitarian organisations are reluctant to use information from social media when responding to crises or conflicts, identifying trust and accuracy as principal concerns. However, the Geographic Information Science literature contains significant research into uncertainty, research we draw upon here to characterise locality descriptions in incident reports related to the 2010 earthquake in Haiti. We do so using a classification developed to georeference locality descriptions in MaNIS, the *Mammal Networked Information System*. We found that although there are similarities between the datasets, crowdsourced crisis information presents significant challenges with respect to vagueness, ambiguity and precision (resolution).

KEYWORDS: Uncertainty, Locality Descriptions, Crowdsourced Crisis Information, Ushahidi

1. Introduction

People affected by crisis or conflict events are turning to social media to communicate with the ‘outside’ and the ‘inside’ world (Coyle and Meier, 2009). On the one hand, humanitarian organisations are reluctant to use information from social media in the response effort (Tapia et al., 2011) because the risks of using untrustworthy and inaccurate information are considerable (Coyle and Meier, 2009). On the other hand, organisations such as Ushahidi have sought to mitigate these risks by developing software to gather, augment and verify crisis information (Ushahidi, 2011c). However, unlike similar organisations such as MapAction (MapAction, 2012), within Ushahidi these tasks are *crowdsourced*, or completed by a heterogeneous group in response to an open call (Howe, 2009).

Accuracy and trust (credibility) are characteristics of uncertainty (MacEachren et al., 2005). Geographic Information Science (GISc) has made considerable progress in evaluating and communicating the uncertainty associated with geographic information (Devillers et al., 2010) and uncertainty is a familiar topic in the GISc literature (MacEachren et al., 2005). Consequently, GISc is well placed to help evaluate the uncertainty associated with crowdsourced crisis information. As a first step towards this evaluation, we consider accuracy. We address two research questions: (1) What types of locality descriptions are present in crowdsourced crisis information? (2) Are the proportions of these types different to those present in related datasets? To do so, we adapt an existing classification of locality descriptions present in MaNIS, the *Mammal Networked Information System*, and apply it to crowdsourced crisis information.

2. Literature review

Several studies have explored the geographic nature of crisis information, especially collections of short text messages (‘microtext’) such as ‘tweets’ related to earthquakes, floods and wildfires (Gelernter and Mushegian, 2011; Vieweg et al., 2010). These studies suggest crisis information contains references to well defined geographic objects, especially when the nature of the event does not imply its location (Vieweg et al., 2010). However, these studies do not attempt to account for the uncertainty associated with these geographic objects.

Where geographic objects are well defined, uncertainty is caused by error (Fisher, 1999). Accuracy is well researched in GISc (Fisher, 1999) and techniques have been developed to evaluate the error associated with point, line and polygon objects (Devillers et al., 2010). However, these techniques involve comparing lower accuracy representations to higher accuracy representations (see Goodchild and Hunter, 1997). Consequently, whilst Haklay (2010) is able to evaluate the accuracy of crowdsourced geographic information by comparing an OpenStreetMap dataset to an Ordnance Survey dataset, it is considerably harder to evaluate the accuracy of crowdsourced crisis information because no higher accuracy representations exist.

Wieczorek et al. (2004) present a solution to the problem of evaluating uncertainty without relying on higher accuracy representations—the ‘point-radius’ georeferencing method. They use this method to georeference records in MaNIS, where the spatial component of each record is a description of the location where the specimen was collected. In addition, the point-radius method has been used to georeference historical search and rescue records (Doherty et al., 2011).

In summary, previous applications of the point-radius method and the geographic nature of crisis information suggest the point-radius method can be applied to crowdsourced crisis information. To assess whether this is the case, and to better understand crowdsourced crisis information, we applied the classification of locality descriptions in the MaNIS dataset to a dataset related to the 2010 earthquake in Haiti (Ushahidi, 2009). However, whilst Wieczorek et al. (2004) and Guo et al. (2008) discuss the categories of locality descriptions in the MaNIS dataset, the categories they identify are slightly different. Consequently, we combined the two classifications to form that shown in Table 1. Table 2 shows a comparison of the three classifications.

Table 1: Combined classification of locality descriptions (following Wieczorek et al., 2004 and Guo et al., 2008)

Code	Category	Example
U	Unsure	
C	Coordinates	
F	Feature	“Springfield”
P	Path	“Hwy. 1”
J	Junction	“Confluence of Labarge Creek and South Labarge Creek”
FOH	Offset from a feature or path at a heading	“10km N of Kuala Lumpur”
NF	Near a feature or path	“Big Bay vicinity”
FS	Subdivision of a feature or path	“N part of Mono Lake”
FOO	Orthogonal offsets from a feature	“1 miles N, 3 miles W of Fairview”
FH	Heading from a feature, no offset	“W of Tucson”
FO	Offset from a feature or path, no heading	“5km outside Calgary”
BF	Between features or paths	“Between Point Reyes and Inverness”

Table 2: Combined classification of locality descriptions compared to Wieczorek et al. (2004) and Guo et al. (2008)

Code	Wieczorek et al. (2004)	Guo et al. (2008)
U	Dubious, Cannot be located, Demonstrably inaccurate	
C	Coordinates	
F	Named place	Feature
P		Path or linear feature
J		Junction
FOH	Offset at a heading	Offset from a feature (or a path) at a heading
NF		Near a feature or a path
FS		Subdivision of a feature or a path
FOO		Orthogonal offsets from a feature
FH		Heading from a feature, no offset
FO	Offset, Offset along a path	Offset from a feature, no heading
BF		Between features or paths

3. Data

The Haiti Crisis Map (Ushahidi, 2009) is an Ushahidi *deployment*—an instance of the Ushahidi software platform—that was set up in response to the 2010 earthquake in Haiti. All 3,606 incident reports that comprise the Haiti Crisis Map were downloaded as a comma-separated values file. Table 3 contains one example.

Table 3: Example incident report from the Haiti Crisis Map (Ushahidi, 2009)

Attribute	Example value
id	3923
title	IDP camp of 250 families has no aid, Cite Soleil
date	2010-03-28 22:00:00
location	Pois Congo, Cite Soleil
description	IDP camp of 250 families in Pois Congo in Cite Soleil ...
category	2b. Penurie d' eau | Water shortage, ...
latitude	18.607433
longitude	-72.319667
approved	YES
verified	YES

Whilst people can report incidents based on their own knowledge or experience, they can also do so based on secondary sources such as SMSs, emails and social media. Consequently, when an incident is reported, several of the attributes in Table 3 may not have values. Typically, one team of volunteers will georeference the 'location' and populate the 'latitude' and 'longitude' attributes (Ushahidi, 2011a), whilst another will approve and verify the incident report (Ushahidi, 2011b). However, incident reports are not versioned, so it is impossible to determine how an incident report changes—and who made those changes—over time.

4. Methodology

The lead author and two additional participants (P1, P2 and P3) independently classified the locality descriptions in the Haiti dataset. Although not experts in the geography of Haiti, all have undergraduate geography degrees, two have postgraduate geographic information systems degrees and all are research students who routinely work with geographic information. In this respect, each participant performed a role that Goodchild (2009) argues is central to academic geography; providing 'quality control' in situations where individuals whose 'activity space' intersects with the study area are unavailable.

To avoid bias, each participant was given a spreadsheet within which row order was randomised and the 'id' attribute was hidden. In addition, each participant was given the information in Table 1 to guide the classification process. In cases where participants were unsure about which category a textual location belonged, they were instructed to select 'Unsure' and comment on their rationale. This captured some of the uncertainty associated with the classification process.

Although time-consuming (it took approximately four hours for each participant to classify the Haiti dataset), a manual classification process has been used in similar research (Gelernter and Mushegian 2011; Vieweg et al., 2010) and captures some of the uncertainty associated with the classification process.

5. Results

For all participants, the most frequent category in the Haiti dataset is 'Feature'. 'Path' is second for P1 and P2, and third for P3; 'Unsure' is second for P3, third for P1 and fifth for P2 (Figure 1). Overall, participants were in agreement in 63.8% of cases (2302), partial agreement in 26.3% of cases (947) and disagreement in 9.9% of cases (357).

To allow a like-for-like comparison between the Haiti and the MaNIS datasets, partial agreement cases were classed by simple majority vote and disagreement cases were classed as 'Uncertain'. All 385 'Uncertain' cases (357 disagreement cases plus 28 'Uncertain' cases) and 19 'Coordinates' cases were then removed. Figure 2 illustrates that in both datasets, the largest proportion of cases are categorised 'F' (51.0% MaNIS, 81.6% Haiti).

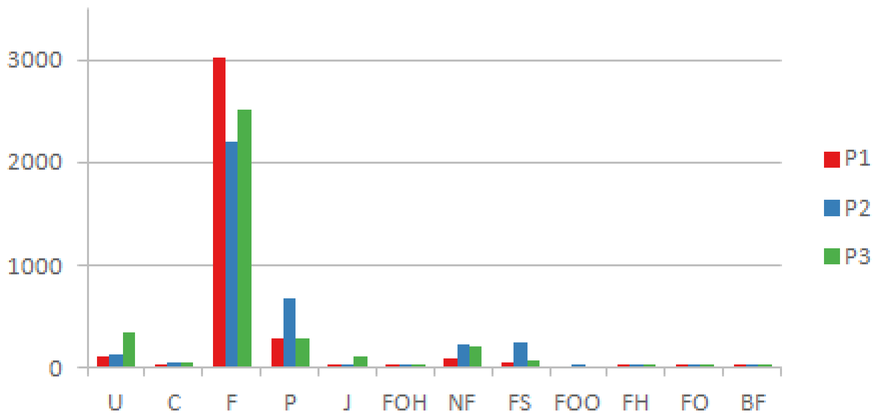

Figure 1: Category frequencies by participant, Haiti dataset

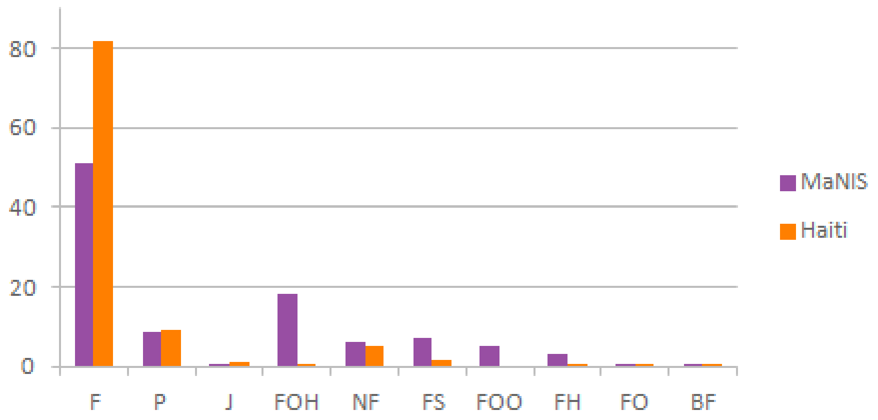

Figure 2: Category distributions, MaNIS and Haiti datasets

6. Discussion

The similarities between the datasets suggest that the point-radius georeferencing method could be applied to the Haiti dataset. However, the results suggest this process would be far from straightforward.

According to Guo et al. (2008), a locality description consists of a target object that may be linked to one or more referenced objects (normally toponyms) by one or more spatial relationships. Implicitly, therefore, a locality description describes a single, unambiguous location. However, participants identified several cases in the Haiti dataset where target objects were ambiguous and referenced objects were vague (for example “Rue Christ-Roi, this is near Hospital Christ-Roi”). Following the instructions, participants classified locality description as ‘Unsure’ and commented on their rationale. However, the ability to evaluate accuracy by exploring differences within, as well as between, locality descriptions requires further analysis. Certainly the vagueness and ambiguity (Fisher, 1999) and precision (resolution) (Veregin, 1999) associated with locality descriptions present interesting research directions.

Although participants attempted to classify locality descriptions consistently, they were uncertain as to whether they did so accurately. Participants related their uncertainty to limited local knowledge: Not being accustomed to the conventions by which, for example, addresses are recorded in Haiti meant they had difficulty distinguishing road names from district names, or road numbers from address numbers. This uncertainty is evident in the 9.9% of cases (357) where participants were in disagreement and questions the assertion that individuals are able to recognise city or street names easily, even when those names are unfamiliar (Gelernter and Mushegian, 2011). However, we argue that such uncertainty is typical in humanitarian response scenarios, especially when the response effort is crowdsourced.

7. Conclusions

This research is a first step towards evaluating the uncertainty associated with crowdsourced crisis information. Results suggest that locality descriptions in the Haiti dataset are predominantly features and that the distribution of locality descriptions across categories is similar to the MaNIS dataset. In turn, this suggests suitable georeferencing methods exist to allow accuracy to be evaluated.

Nevertheless, this conclusion is partial and hides the complexities present in crowdsourced crisis information. To address these complexities we plan to investigate whether alternative sources of information such as OpenStreetMap can be used to overcome limited local knowledge and explore differences within locality descriptions. We also plan to extend our research to a similar dataset related to the recent conflict in Libya (OCHA, 2011).

8. Acknowledgements

We thank Roger Beecham and Sarah Goodwin for classifying the Haiti dataset and those behind the Haiti Crisis Map for making their data publicly available. The bars in Figures 1 and 2 are elements A to E in ColorBrewer Set 1 (Harrower and Brewer, 2003). This qualitative scheme is well suited to categorical data. These five elements are effective on screen and in print.

9. References

Coyle, D. and Meier, P., 2009. *New Technologies in Emergencies and Conflicts: The Role of Information and Social Networks*, Washington DC and London: UN Foundation-Vodafone Foundation Partnership. Available at: http://www.unfoundation.org/press-center/publications/new-technologies-emergencies-conflicts.html [Accessed June 2, 2011].

Devillers, R., Stein, A., Bédard, Y., Chrisman, N., Fisher, P. and Wenzhong, S., 2010. Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities. *Transactions in GIS*, 14(4), pp.387-400.

Doherty, P., Guo, Q., Liu, Y., Wieczorek, J. and Doke, J., 2011. Georeferencing Incidents from Locality Descriptions and its Applications: a Case Study from Yosemite National Park Search and Rescue. *Transactions in GIS*, 15(6), pp.775-793.

Fisher, P., 1999. Models of uncertainty in spatial data. In P.A. Longley, M.F. Goodchild, D.J. Maguire, D.W. Rhind and P. Fisher, eds. *Geographical Information Systems, Principles and Technical Issues*. Chichester: John Wiley and Sons, pp. 191-205.

Gelernter, J. and Mushegian, N., 2011. Geo-parsing Messages from Microtext. *Transactions in GIS*, 15(6), pp.753-773.

Goodchild, M., 2009. NeoGeography and the nature of geographic expertise. *Journal of Location Based Services*, 3(2), pp.82-96.

Goodchild, M.F. and Hunter, G.J., 1997. A simple positional accuracy measure for linear features. *International Journal of Geographical Information Science*, 11(3), pp.299-306.

Guo, Q., Liu, Y. and Wieczorek, J., 2008. Georeferencing locality descriptions and computing associated uncertainty using a probabilistic approach. *International Journal of Geographical Information Science*, 22(10), pp.1067-1090.

Haklay, M., 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design*, 37(4), pp.682-703.

Harrower, M. and Brewer, C., 2003. ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. *The Cartographic Journal*, 40(1), pp.27-37.

Howe, J., 2009. *Crowdsourcing: How the Power of the Crowd is Driving the Future of Business*, London: Random House Business Books.

MacEachren, A.M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M. and Hetzler, E., 2005. Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. *Cartography and Geographic Information Science*, 32(3), pp.139-160.

MapAction, 2012. MapAction: About us. Available at: http://www.mapaction.org/about/about-us.html [Accessed February 8, 2012].

OCHA, 2011. Libya Crisis Map. Available at: http://libyacrisismap.net/ [Accessed October 18, 2011].

Tapia, A.H., Bajpai, K., Jansen, J., Yen, J. and Giles, L., 2011. 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 Information Systems for Crisis Response and Management (ISCRAM)*. The 8th International Conference on Information Systems for Crisis Response and Management (ISCRAM). Lisbon, Portugal.

Ushahidi, 2009. Haiti Crisis Map. Available at: http://haiti.ushahidi.com/ [Accessed October 18, 2011].

Ushahidi, 2011a. How To Map SMS Messages. Available at: http://community.ushahidi.com/uploads/documents/c_How-To-Approve-SMS.pdf [Accessed October 19, 2011].

Ushahidi, 2011b. Ushahidi Guide To Verification. Available at: http://community.ushahidi.com/uploads/documents/c_Ushahidi-Verification-Guide.pdf [Accessed October 20, 2011].

Ushahidi, 2011c. Ushahidi: About us. Available at: http://www.ushahidi.com/about-us [Accessed October 19, 2011].

Veregin, H., 1999. Data quality parameters. In P.A. Longley, M.F. Goodchild, D.J. Maguire, D.W. Rhind and P. Fisher, eds. *Geographical Information Systems, Principles and Technical Issues*. Chichester, United Kingdom: John Wiley and Sons, pp. 177-189.

Vieweg, S., Hughes, A.L., Starbird, K. and Palen, L., 2010. Microblogging during two natural hazards events: what Twitter may contribute to situational awareness. In *Proceedings of the 28th international conference on human factors in computing systems*. The 28th international conference on human factors in computing systems. Atlanta, GA, pp. 1079–1088.

Wieczorek, J., Guo, Q. and Hijmans, R., 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science*, 18(8), pp.745-767.

10. Biography

Iain Dillingham is a research student at the giCentre, City University London, applying visual and analytical approaches to user-generated content gathered in the wake of crisis or conflict events.

Dr Jason Dykes is Professor of Visualization at the giCentre, City University London, undertaking applied and theoretical research in, around and between information visualization, interactive analytical cartography and human-centred design.

Dr Jo Wood is a Professor of Visual Analytics at the giCentre, City University London, with research interests in geovisualization and visual analytics of data with a temporal and spatial component.