A Model for Assigning Uncertainty to Non-Participatory Volunteered Geographic Information

Ian Kramer, University of South Carolina

The evolution of Internet and mobile technology has made it increasingly easy for the general public to contribute vast amounts of information to a globally connected community. The introduction of Web 2.0 technology and smart phones has provided a mechanism for people to submit various forms of spatially referenced data via the Internet. Various groups are interested in using this data to understand social trends that may exist through space and time. Geographers are especially interested in the spatial aspect of this public commentary to understand patterns in various aspects of social behavior. One of these geographic research initiatives looks at leveraging this globally connected community to provide data to support emergency response efforts during disaster events.

The term "Volunteered Geographic Information" or VGI has been coined to describe the generation of spatial data and maps by the general public (Goodchild, 2007). The majority of VGI is created through a participatory effort that provides geographic data for social, commercial, political, and disaster related purposes. The initial concept of VGI focused on an individual or group of people creating spatial data sets to support one of the aforementioned purposes. This capability gave the public the ability to create maps, thus starting a trend of moving map production away from the centralized production model that was used in the past (Goodchild, 2007).

Recent trends in Internet use have also seen the rise of social networking. Social networking provides the ability for a person to share personal data with other people through a socially-linked network (Boyd and Ellison, 2007). There has been a 174% increase in the size of the social networking audience from 2007-2011 (Shaw, 2012). Due to the large amounts of participants, as well as commentary present in social networking, various groups are interested in mining the data to understand social, commercial, and political trends over a specific temporal space. Data derived from social networks can be considered a form of VGI data, because it is still volunteered. However the primary motivation behind its creation typically does not include the creation of spatial data or the production of maps. Because of its non-participatory role in the creation of spatial data, this research will refer to this subset of VGI data as Non-Participatory Volunteered Geographic Information (NPVGI).

Recently, there has been increased interest in using VGI and NPVGI data in the area of situational awareness. The research has focused on analyzing the role of VGI and NPVGI in the emergency response for earthquakes (Guy et al., 2010; Starbird, 2011), flooding (Poser and Dransch, 2010), fire (De Longueville et al, 2010), and public health (Kamal Boulos et al., 2010). The use of NPVGI provides a source of information to responders which allow them to collect commentary, audio, pictures, and video from the field without having to deploy valuable resources. VGI also provides a mechanism to increase mapping resources during an event using a distributed network of volunteers (Goodchild and Glennon, 2010). The scope of this research will focus on the use of NPVGI within the realm of situational awareness and disaster response.

The use of NPVGI in situational awareness inherits some of the current limitations with VGI data as a whole. One of the primary limitations of NPVGI is the lack of ownership and metadata which helps establish the credibility of the data (Goodchild, 2007; Goodchild and Glennon, 2010). The absence of metadata provides a high level of uncertainty about the data, which reduces a responder's overall confidence in the data set (Goodchild and Glennon, 2010). This lack of information introduces questions about the contributor and their motivation behind the submission. Due to this unknown uncertainty, responders do not use NPVGI to its fullest potential during events.

Current research is looking to improve the identification of a NPVGI data set's location. A large portion of NPVGI data is geo-referenced incorrectly or lacks spatial coordinates (MacEachren et al., 2011). In the absence of locational information, recent methodologies analyze the context of the NPVGI data to identify references to place names (Guy et al., 2010; MacEachren et al., 2011). A place name can then be geo-referenced against a database of know places to provide an estimated spatial location. This methodology introduces uncertainty into the placement of the NPVGI data because it relies on the assumption that the context accurately describes its location.

The use of Natural Language Processing has been proposed to classify the context of NPVGI (Verma et al., 2011). This methodology identifies the tone and the type of linguistic register used in the context of the NPVGI data set. The results of this methodology provide a contextual classification schema that defines NPVGI has objective or subjective. This approach could be further expanded to assess the accuracy of the contextual and locational properties of the NPVGI data set.

The limitations of using NPVGI data contribute to its overall uncertainty. There has been little research done in addressing the credibility and uncertainty of NPVGI data. The current research on VGI and NPVGI have identified uncertainty as one of the main future research objectives (Mummidi and Krum, 2008; Flanagin and Metzger, 2008; Goodchild and Glennon, 2010; Longueville et al., 2010; Osterman and Spinsanti, 2011; Guy et al., 2010; Poser and Dransch, 2010). Osterman and Spinsanti (2011) have proposed a generalized approach for assessing credibility for NPVGI data. However, this approach was never implemented or tested against NPVGI data sets.

The proposed dissertation research aims to study the uncertainty of NPVGI and how it can be applied to situational awareness. This research will build on existing methodologies used to determine geographic location for NPVGI data that is collected from various platforms. A key goal in this research is to develop a credibility and uncertainty rating system which allows users to understand the potential error associated with the data. An uncertainty rating system will provide a better method for assessing the quality of NPVGI data for decision makers.

This research first proposes to develop a model to determine the geographic location of NPVGI data. This portion of the research will expand on methodologies proposed by Guy et al. (2010) and MacEachren et al. (2011). The goal of this portion of the research is to show that temporal-contextual information can be used to identify a location of a NPVGI data set that does not contain a reference to a place name. This process will look for contextual correlations with known locations of NPVGI data that exist within the same temporal window.

The second part of this research proposes to develop a model to assess the credibility of NPVGI based on the spatio-temporal relationship with neighboring NPVGI data and ancillary data sources. This part of the research will improve the approach suggested by Osterman and Spinsanti (2011) by integrating Natural Language Processing into the credibility assessment model. This model will be used to assign a credibility value to the NPVGI data as well as the content provider.

The third part of the research includes developing a model to identify uncertainty in the NPVGI collection process. This portion of the research will use the uncertainty classifications proposed by MacEachren et al. (2004). However, this research will only focus on five of the classifications which include: precision, credibility, interrelatedness, subjective and accuracy uncertainty. This model will be used to assign individual uncertainty values for each category and a combined uncertainty score.

The use of NPVGI data during a disaster event can provide responders with a large amount of information in the form of real-time commentary, audio, photos, and video from the affected area. The ability to quantify the credibility and uncertainty of the information will increase the confidence in the validity of the content which ultimately will further promote its use. The proposed model can also be applied to other uses outside the realm of disaster response including academic, commercial, governmental, and media uses.

References Cited

Boyd, D.M., Ellison, N.B., 2007. Social Network Sites: Definition, History, and Scholarship, Journal *of Computer-Mediated Communication*, 13(1):210-230.

De Longueville, B., Annoni, A., Schade, S., Ostlaender, N., Whitmore, C., 2010. Digital Earth's Nervous System for crisis events: real-time Sensor Web Enablement of Volunteered Geographic Information, International *Journal of Digital Earth*. 3(3):242-259.

Flanagin, A.J., Metzger, M.J., 2008. The credibility of volunteered geographic information, GeoJournal, 72:137-148.

Goodchild, M.F., 2007. Citizens as sensors: the world of volunteered geography, GeoJournal. 69:211-221.

Goodchild, M.F., Glennon J.A. 2010. Crowdsourcing geographic information for disaster response: a research frontier, *International Journal of Digital Earth*, 3(3):231-241.

Guy M., Earle, P., Ostrum, C., Gruchalla, K., Horvath, S., 2010. Integration and Dissemination of Citizen Reported and Seismically Derived Earthquake Information via Social Network Technologies, Advances *in Intelligent Data Analysis IX, Lecture Notes in Computer Science*, 6065:42-53.

Kamel Boulus, M.N., Sanfilippo, A.P., Corley, C.D., Wheeler, S., 2010. Social Web mining and exploitation for serious applications: Technosocial Predictive Analysis and related technologies for public health, environmental, and national security surveillance, *Computer Methods and Programs in Biomedicine*, 100:16-23.

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

MacEachren, A.M., Jaiswal, A., Robinson, A.C., Pezanowski, S., Savelyev, A., Mitra, P., Zhang, X., Blanford, J., "SensePlace2: GeoTwitter analytics support for situational awareness," *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference*, 23-28 Oct. 2011, pp.181-190.

Mummidi, L.N., Krumm, J., 2008. Discovering points of interest from users' map annontations, *GeoJournal*, 72:215-227.

Ostermann, F.O., Spinsanti, L., 2011. A Conceptual Workflow for Automatically Assessing the Quality of Volunteered Geographic Information for Crisis Management, *AGILE 2011*, 18-22 April, 2011.

Poser, K., Dransch, D., 2010. Volunteered geographic information for disaster management with application to rapid flood damage estimation, *Geomatica*, 64(1):89-98.

Shaw, M. 2012. The State of Social Media, Online Marketing Institute, London, *OMI Social Media Summit*. URL: http://www.comscore.com/Press_Events/Presentations_Whitepapers/2012/The_State_of_Social_Media (Last date accessed: 30 April, 2012).

Starbird, K., 2011. Digital Volunteerism During Disaster: Crowdsourcing Information Processing, *CHI 2011 Workshop on Crowdsourcing and Human Computation*, 7-12 May, 2011, Vancouver, BC, Canada.

Verma, S., Vieweg, S., Corvey, W.J., Palen, L., Martin, J.H., Palmer, M., Schram, A., Anderson, K.M., 2011. Natural Language Processing to the Rescue? Extracting "Situational Awareness" Tweets During Mass Emergency, *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, *ICWSM 2011*, 17-21 July, 2011, Barcelona, Spain.