Supporting the Geolocating Work of Crisis Mappers

Kate Starbird
Department of Human Centered Design & Engineering
University of Washington
kstarbi@uw.edu

ABSTRACT
This paper discusses challenges of geolocating social media updates during crises, looking to the work of crisis mapping volunteers both to inform future computational solutions and as a design opportunity for GeoHCI researchers.

Author Keywords
Crisis Informatics; crowdsourcing; crisis mapping; location; geolocation; geography

ACM Classification Keywords
H.5.3 Groups & Organization Interfaces—collaborative computing, computer-supported cooperative work

INTRODUCTION

I did not intend to become a GeoHCI researcher. For me, there is nothing inherently interesting about geography, but there is something inherently geographical about my research area: crises. My work focuses broadly on the use of social media (SM) during natural disasters and other mass disruption events. The underlying hypothesis driving this research is that people experiencing crisis events are newly enabled to share information with each other and with the global audience, and that this information has the potential to improve *situational awareness* and help affected people and crisis responders make better-informed decisions [5]. Initially, I went searching for solutions to help make sense of what has become a flood of data flowing through these platforms during large-scale crisis events. That search led me to the *digital volunteer* communities working to assist during events by manually processing information from SM and other sources [2,3]. And those volunteers led me to the maps.

Crisis Mapping Communities

Several crisis mapping communities have sprung up in recent years. Often, these groups have emerged in response to a specific event. In some cases they have formalized into ongoing organizations. CrisisMappers, the Standby Task Force (SBTF), and the OpenStreetMap Humanitarian Operations Team (HOT) are all digital volunteer communities focused on creating maps during crisis events. While the latter group, heralded for creating the most usable map for responders during the 2010 Haiti earthquake response, concentrates on mapping static geographical information, the former groups work to map evolving humanitarian conditions during crisis events: Where are people? What do they need? These groups often rely on SM updates and other citizen reported information as primary sources. Volunteers, some trained and some not, work to identify actionable or otherwise useful information, verify it, and then to find GPS coordinates for it so they can add it to their public maps. Examining this crisis mapping activity reveals opportunities for GeoHCI researchers to contribute to solutions that help crisis responders and those affected by disasters by making the information shared on SM usable.

GeoHCI Workshop at CHI 2013, April 27–28, 2013, Paris, France

Tweak the Tweet: A Resource for Digital Volunteers

In 2009, we proposed Tweak the Tweet (TtT), a crisis reporting microsyntax that enables Twitter users to create tweets that are essentially machine-readable [4]. Though the syntax has proved hard for affected people to use, it has often been adopted by digital volunteers [2]. Assisted by my colleagues at the University of Colorado along with hundreds of digital volunteers, I have deployed TtT for more than 30 events since 2010. During these events, I provide a Google Map that displays TtT tweets automatically geolocated, primarily using location information in the text, by software built to support TtT. Many insights conveyed in this paper are lessons learned from efforts to support TtT and from interviewing digital volunteers who use the syntax.

CHALLENGE: EXTRACTING PRECISE AND ACCURATE LOCATION FROM SOCIAL MEDIA UPDATES

For SM information to be useful in the crisis context, it must be geolocated with both accuracy and precision. Extensive research on digital volunteer communities, including an extended period as a participant-observer within Humanity Road, a virtual organization that often participates in crisis mapping efforts [3], along with lessons learned from more than 30 deployments of TtT, has exposed some of the complexity of this problem as well as some strategies for solving it. Superficially, there are two methods of geolocating SM updates: 1) using geolocation information imbedded in the metadata of messages; and 2) extracting location from the textual content of messages.

Using Geolocation Information in Metadata of Updates

Many SM platforms allow users to automatically add GPS coordinates to the metadata of their updates. This method of geolocating, addressed through user preferences and actions within SM client applications, is perhaps ideal for reporting information from the ground of crisis events. However, currently only a small percentage of SM updates have this geolocation information in their metadata. For example, on Twitter, perhaps the most popular platform for crisis data reporting, only 0.8% of messages about the 2012 Colorado fires contained GPS coordinates. We found, through TtT deployment efforts, that Twitterers had a hard time activating geolocation, especially through their mobile devices where they had to change preferences in two separate locations. However, other SM sites have a higher percentage of geolocated updates. During the Colorado fires event, 8% of Instagram photos that were also posted to Twitter had geolocation information in their metadata. This suggests, not surprisingly, that the design of SM services and the client applications that access them affect the likelihood of updates being geo-stamped. It also suggests Instagram may be an important, usable source of info coming from the ground of crisis events in the future, though recent changes that complicate its connection to Twitter [1] may dampen its value in this context.

Another problem with automatic geolocation of SM updates involves the posting of secondhand information. This can occur when an affected person posts a message about a location after they have evacuated; when an account owner acts as a proxy for another person, posting on their behalf; or when remote volunteers post information about places where they are not. For TtT, we found that we needed to preference textual location (where present) ahead of metadata location, because TtT tweets were far more likely to be composed by someone outside the affected area.

Extracting Geolocation from Textual Content

Extracting geolocation from the textual content of SM updates solves many of the issues with automatic geotagging. It doesn't require individuals to turn on geolocating services, and it alleviates the secondhand problem, as it geolocates the area referred to in the message, not where the message came from. However, even leveraging the power of existing libraries for matching text to location, there are several challenges with this strategy.

One major issue is disambiguating location information. For events like hurricanes that strike a large area, this can be especially vexing. Every city has a Main Street. Every coastal state has a Newport. We often confront this problem during TtT deployments. For example, when a tropical storm hit Louisiana in 2011, we struggled to differentiate between street names in New Orleans (e.g. St. Charles) and parishes by the same names, which were often miles away. Disambiguation work had to be done manually.

Another difficulty is identifying where location information is within text. TtT solves this with human computation, by asking Twitterers to mark tweets with a #loc tag to designate where location information is within the tweet. Automatic techniques may be able to locate city names or neighborhoods, but may not be able to pick up more subtle pieces of location information, lowering precision.

LEVERAGING THE WORK OF DIGITAL VOLUNTEERS

Research on crisis mappers and other digital volunteer communities shows the state of the art in geolocating SM updates to rely heavily on human computation. Members of these communities use a variety of different strategies to find actionable information, verify it, and get it onto their maps. Studying how these volunteers work may illuminate possible solutions for enhancing our collective ability to map this potentially valuable information. Future research should look towards developing better computational strategies for automatically geolocating SM updates as well as designing tools to help digital volunteers and other *crowdworkers* better do this work themselves.

Previous research on crisis mapping work suggests that the best approach may be a holistic one, combining information from textual content in a current SM message with that account's history including any updates that have geolocation metadata, along with profile location, friend networks, etc. Volunteers mapping Twitter information currently do this work manually, reading through tweets and profiles, combining street address location in tweets with city and neighborhood location in profiles, and reading tweet histories where available to determine if the source is actually on the ground in the area. We can learn from these strategies to design tools to automate some of this work. However, complete automation of the process may not be desirable, as the difficulty of disambiguation and the safety-critical nature of the domain mean that even highly accurate machine learning techniques (e.g. 90%) may not be good enough. We should therefore also continue to design to support crisis mapping volunteer work.

REFERENCES

1. Lunden, Ingrid. What the Twitter/Instagram Standoff Has Meant for Traffic to Instagram. *TechCrunch*, December 17, 2012.
2. Starbird, K. & Palen, L. "Voluntweeters:" Self-Organizing by Digital Volunteers in Times of Crisis. *Proc. of CHI 2011*, 1071–1080.
3. Starbird, K. & Palen, L. Working & Sustaining the Virtual 'Disaster Desk.' *Proc. of CSCW 2013.* Forthcoming.
4. Starbird, K. & Stamberger, J. Tweak the Tweet: Leveraging Microblogging Proliferation with a Prescriptive Syntax to Support Citizen Reporting. *Proc. of ISCRAM, 2010*.
5. Vieweg, S, Hughes, A, Starbird, K, & Palen, L. Micro-Blogging during Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness. *Proc of CHI 2010*, 1079-88.