Position Paper: Extracting Geospatial Information from Social Media

James M. Kang, Ashley Holt, John Greer
National Geospatial-Intelligence Agency
7500 GEOINT Drive, Springfield VA 22150

The past decade has seen the sudden attention to and assimilation of social media technologies all over the world. Social media (e.g., Twitter, Facebook) has been referenced as one of the major driving forces of the recent Arab Spring spreading across the middle east (e.g., see Stefanidis et al. 2011). The immense amount of data generated from social media provides opportunities for researchers to understand and extract useful information about social networks, cultures, sentiments, and the transmission of ideas through populations.

In recent years, social media technologies (e.g., Twitter and Facebook) are also beginning to embrace the usefulness of having a spatial reference along with the messages. Combining Geospatial information within social media may identify not only location, but also local culture, "slang" terms and place names unique to that location.

Utilizing geospatial location and natural language processing techniques may help in extracting some spatial information. For example, Cheng et al. utilized geospatially referenced Twitter information to geo-locate messages based on content [Cheng 2010]. Building on this work, Cano et al. leveraged temporal information in social media for identifying relationships between geographic location and the social constructs associated with place, which may change over time.

There exist several main challenges when extracting geospatial information from social media; for example the user-contributed information may not have any geospatial identifiable text (e.g., geo-coordinate) within it. In addition, the massive amount of social media produced daily causes significant computational challenges to process the datasets. Extracting and processing geospatial information from these massive datasets in near real-time requires state-of-the-art large-scale computing methods, including cloud computing.

References

[Cano 2011] Cano, A.E., A. Varga, and F. Ciravegna (2011). Volatile Classification of Point of Interests based on Social Activity Streams. Procedings of the 10th International Semantic Web Conference, Bonn, Germany, October 23-27, 2011.

[Cheng 2010] Cheng, Z., Caverlee, J., and Lee, K. (2010) You are Where You Tweet: A Content-Based Approach to Geo-locating Twitter Users. CIKM'10 Conference Procedings, October 26–30, 2010.

[Stefanidis 2011] Stefanidis, T., A. Crooks, and J. Radzikowski (2011). Harvesting ambient geospatial information from social media feeds. GeoJournal, DOI 10.1007/s10708-011-9438-2.