REMOTELY ASSESSING LOCAL POPULATIONS THROUGH SPATIAL MEDIA ANALYSIS

US Army Corps of Engineers

Mr. Kalev Leetaru
University of Illinois
Dr. Charles R. Ehlschlaeger and Mr. Jeffrey A. Burkhalter
United States Army Engineering Research and Development Laboratories, Champaign, IL
Ms. Marina V. Drigo
PERTAN Inc., Champaign, IL

POC: Dr. Charles R. Ehlschlaeger, Charles.R.Ehlschlaeger@usace.army.mil

BACKGROUND

This research is funded by the US Army Corp of Engineers Engineering Research & Development Center (ERDC). For two years, ERDC has been collaborating with researchers at the 95th Civil Affairs Brigade (CAB) to better understand population attitudes about potential infrastructure and essential service (IES) improvements. CABs, in collaboration with partner nations' military and local governments, prescribe IES improvements to improve governance to regions overrun with criminal organizations. CAB analysts must choose locations for IES improvements as well as the types of IES improvements. The goal of this research is to monitor the population's attitude regarding economic, educational, health, or governmental IES improvements. The final product will be a monitoring tool that will allow CAB teams to combine their expert knowledge of village and county attitudes about IES improvements to the media analysis maps. CAB teams and analysts will then better estimate village and county attitudes in places not directly or recently observed.

ABSTRACT

The systematic assessment of populations through ground-based censuses and surveys has been a mainstay of governmental function for at least five thousand years, collecting critical demographic and economic information on residents and citizenry. As the United States has engaged in nation building and nation strengthening efforts across the world, it has increasingly needed to assess populations in regions that cannot be directly accessed due to cost, security, time, or other restrictions. In particular, the dreams and fears of those populations are of key importance in ensuring that efforts most directly address the needs of those served. Since World War II, Western governments have monitored the media to remotely assess the messages, beliefs, and available information from distant populations, an effort known as Open Source Intelligence (OSINT). The rise of electronic media and increasing computational power has led to a revolution in automated assessment of populations. Ongoing research is exploring a hybrid approach of conflating OSINT, traditional ground-based surveys, and demographic analysis to better understand subpopulation attributes and their beliefs and opinions. This poster will present preliminary results from a rigorous quantitative comparison of ground-based, media-derived, and hybrid assessments of remote populations in several selected regions.

THE RISE OF THE MEDIA DELUGE

In 2012, every day a quarter-billion photographs are uploaded to Facebook, 300 billion emails are sent and 340 million tweets are posted to Twitter. There are more than 644 million websites with 150,000 new ones added each day, and upwards of 156 million blogs. Outside of Asia, Facebook connects more than half of all global Internet users, while in China more than 30% of all Internet users make use of Sina Weibo. Every day, there are 2.5 times as many words posted to Twitter (more than 8 billion) as there were in every article of every issue of the New York Times over the last half-century. By 2015, the 50 trillion Twitter posts will exceed all the words in all the books published over the last half-millennia. With nearly a third of planet's population now connected to the internet, ordinary citizens are becoming a vast "always on" social sensor network, providing a continuous real-time stream of ordinary life and extraordinary events from every corner of the globe. With the number of cell phones now approximately equal to the number of people on earth, bystanders and participants are often the first to report on emerging events, streaming reports, photographs, videos, narratives, and opinions within minutes.

ASSESSING POPULATIONS THROUGH THE MEDIA

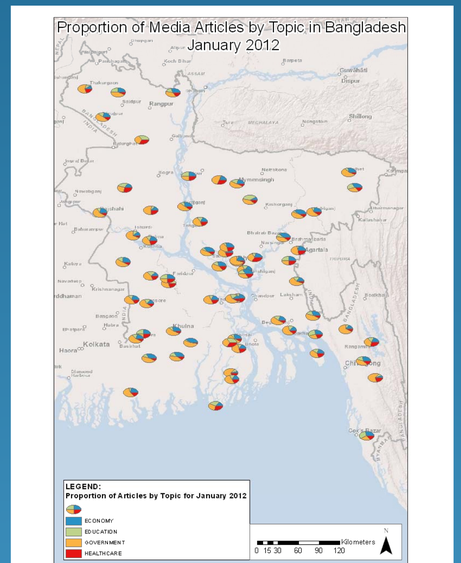
The informational equivalent of all the words ever spoken by humans since the formation of language transits the global internet every 7 days. More importantly, a growing body of work is providing validation that the results of analyzing the global media match reality (the tone of tweets towards political candidates match phone polls while the tone of Twitter as a whole matches circadian rhythms) and most excitingly offer forecasts of future human behavior (Twitter tone forecasts the stock market, while global media tone forecasts country stability). This is in keeping with more than 70 years of open source intelligence study of the media as a mechanism to assess the public information environment, and, increasingly, the views of ordinary citizens.

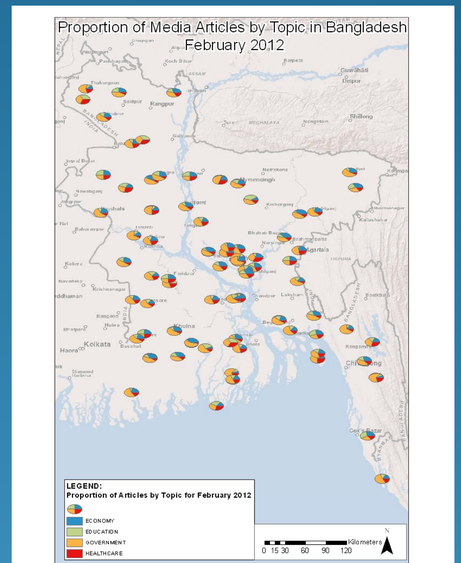
A HYBRID APPROACH

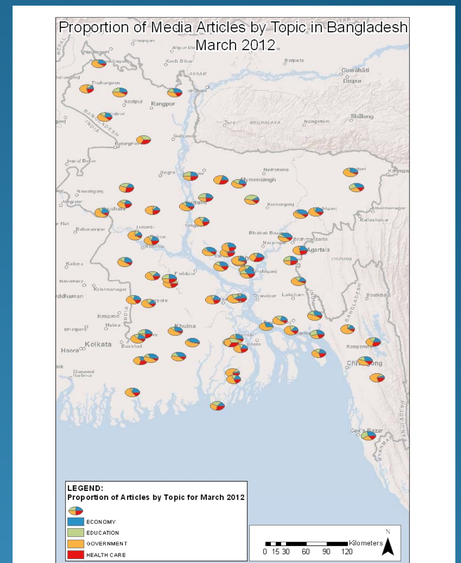
The goal of this research is to identify strategies to find patterns of media tone at large geographic scales in developing nations by exploiting demographic analysis to identify subpopulations. This paper will present preliminary findings from a small selection of geographic regions in which augmented traditional demographic data has been fused with geographically-referenced media-derived population surveys to provide the first quantitative assessment insights into this emerging field.

The system used for this analysis is currently configured to monitor nearly 1,500 major news websites and aggregators around the world representing a cross-section of global events and opinions, including news aggregators like Google News and Yahoo News, international outlets like CNN, BBC, and the New York Times, wire services like AP and Xinhua, regional press like the Times of India and Pakistan's The News and local press like the Jerusalem Post and the Jamaica Gleaner. The system can monitor any data source available in an RSS feed and is easily extensible to monitor other OSINT data sources such as the Open Source Center. New RSS feeds may be added at any time simply by adding it to the monitoring list.

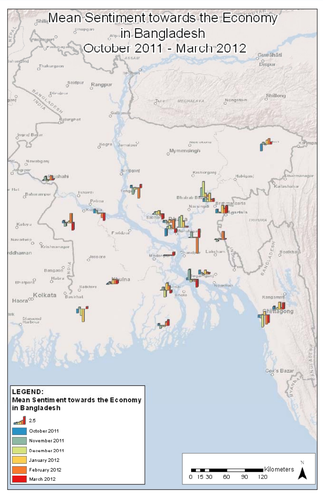
This first set of maps indicates how often IES topics are discussed in social media. Each map represents a month of messages (January 2012 to March 2012) showing the proportion of messages for each of the IES topics (The Economy, Educational Improvements, Governance Improvements, and Health Care Improvements). These maps are currently NOT calibrated to CAB analysts' and teams' understanding of the local conditions. Temporal trends can be discovered by identifying changes in topic discussion frequency that will determine the most effective prescription improvements.

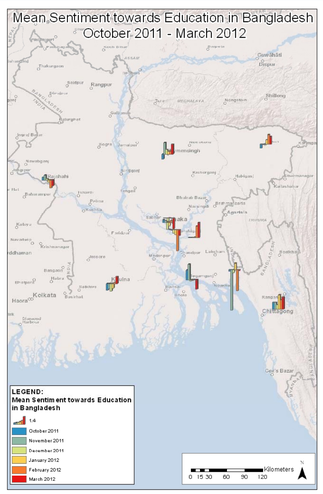


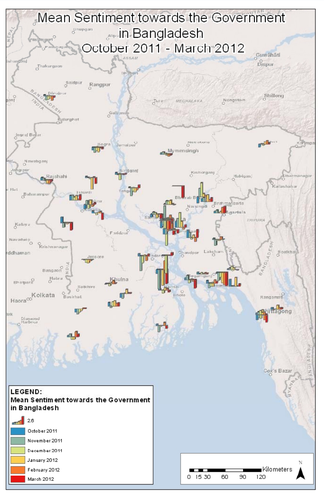


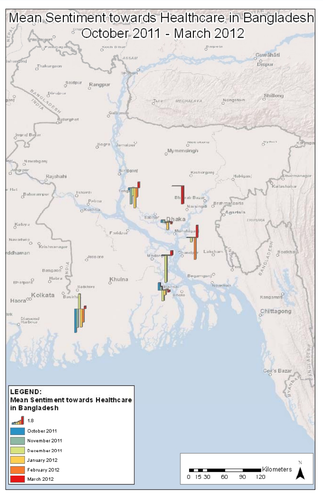


This second set of maps indicates message sediment for IES topics. Each chart bar represents a month of messages (October 2011 to March 2012) showing the magnitude of positive (above the line) or negative statements in the messages for each topic. These maps are currently NOT calibrated to CAB analysts' and teams' understanding of the local conditions. Temporal trends at locations where IES prescriptions were performed can detect the effectiveness of those prescriptions. When compared to the database of natural disasters, terrorist activities, and other drivers of conflict, temporal trends can identify subpopulations working more closely with partner nations or criminal organizations.









BUILDING STRONG®