

# Correlating Twitter Sentiment and Event Data to Monitor Social and Political Unrest

Lucas Overbey, Scott Batson, Jamie Lyle, Christopher Williams, Robert Regal and Lakeisha Williams

SPAWAR Systems Center Atlantic

lucas.overbey@navy.mil, scott.batson@navy.mil,  
jamie.lyle@navy.mil, christopher.s.willli6@navy.mil,  
robert.regal@navy.mil, lakeisha.d.williams1@navy.mil

The readily observable communications found on internet social media sites now play a prominent role in the spread of information which, when accompanied by subjective statements, can indicate public sentiment and perception. In particular, recent international events, including natural disasters, the continuous cascade of protests and revolutions in the Arab world, and terror attacks, have uncovered the utility of social networking sites like Twitter for understanding social and political unrest. Effectively monitoring changes in sentiment toward political events within these regions may reveal trends in public opinion that indicate social and political instability.

Feelings of assurance or stability are more likely to be associated with a content population, while feelings of uncertainty and anger are more conducive to invoking political change. General public sentiment within a region can change with events that affect its population. Understanding what political events are correlated with or influence public sentiment is key for understanding how to respond to events that take place. Furthermore, developing a measure to monitor the emotions that precede social and political revolutions would be helpful in preparing for political instability and understanding the fluctuating allegiances of a volatile population.

Open source automated political event coding schemes and datasets provide information about the interactions between state and non-state actors. For example, GDELT, KEDS, ICEWS, and Phoenix provide data to study international political conflict through processing news stories in traditional media. Phoenix data relies on the Open Source EL:DIABLO project, a process which includes a web-scraping to pull news stories from online Rich Site Summary (RSS) feeds and an automated event coder. PETRARCH, the automated coding library, utilizes natural language processing (NLP), a lexical references consisting of event and actor ontologies and verb and noun phrase dictionaries, and the Conflict and Mediation Event Observations (CAMEO) coding scheme, to output data that includes a source and target actor, an event that took place between the actors, a location, and the date of this event. These datasets have been studied to forecast political instability, identify trends such as escalation of conflict between actors, and monitor interaction between countries.

A key component to understanding public opinion is extraction of the aspect toward which a given sentiment is directed. Given that Twitter is a micro-blogging service that limits messages to 140 characters in length, these aspects are often hidden in links to other sources of information. Twitter users commonly share their opinion

on events described in linked news stories that they find interesting. A metric to quantify the sentiment that social media users express towards political actors and/or politically-relevant events would be useful to understand how these actors or events influence affected populations. Correlating the sentiment of a population with the political events affecting them would provide a near-real-time analysis of how these events impact human behavior and actions. These measures may be used as a variable for predicting political instability, for identifying which events generate the most active responses, and for monitoring social and political unrest. A spatio-temporal analysis of these correlations would even allow one to address questions such as whether specific political events, or gradual shifts in sentiment towards political actors, serve as a greater catalyst for societal and policy changes. The goal of our work is to construct a method to monitor the shifts in sentiment towards political actors that correspond to political events by correlating Twitter messages with Phoenix event data. This approach also provides a metric for better understanding the political stability of a geographic region, and will give insight into which events resonate within a population.

To develop a methodology for correlating social media and political event data, we use a self-collected Twitter dataset. We use broad search terms related to European governmental organizations and known events to collect Twitter data. Use of broad terms around governmental social media accounts allows us to collect a large social media dataset that will likely contain a substantial number of Uniform Resource Locator (URL) links to politically-relevant news stories.

From the collected Twitter data, we identify tweets that contain URL links and follow the links to sources regarded as trusted news outlets. We then use the Phoenix Pipeline (the EL:DIABLO/PETRARCH system) to extract geo-political events using automated coding. Once an event is identified as the subject of a tweet, we then correlate the event type and Goldstein Score with a sentiment analysis on the text in the original and related tweets. We also compare sentiment within certain geographic regions to actors within and outside of these regions. In addition, we perform Latent Semantic Analysis (LSA) as a form of topic modeling within the Twitter dataset. We then analyze this Twitter message content-based topic modeling in relation to the linked event codes.

Our approach introduces a novel strategy to extract aspect in sentiment analysis by using link-following and automated event coding. Results indicate a correlation between news event codes and public sentiment expressed toward these events on Twitter. Our study yields insight into how Twitter users respond to an event. One may identify trends in this correlation over time to study and predict social and political changes. In addition to being a novel analysis for monitoring general changes in sentiment towards political actors and/or events, sampling social media users for this correlation could also yield a novel approach to quantifying the stability of a population within some geographic region of interest.