Linking Twitter Sentiment and Event Data to Monitor Public Opinion of Geopolitical Developments and Trends

Lucas A. Overbey, Scott C. Batson, Jamie Lyle, Chris Williams, Robert Regal, and LaKeisha Williams

Space and Naval Warfare Systems Center Atlantic
P.O. Box 190022
North Charleston, SC 29419-9022
lucas.overbey@navy.mil

Abstract. Readily observable communications found on Internet social media sites can play a prominent role in spreading information which, when accompanied by subjective statements, can indicate public sentiment and perception. A key component to understanding public opinion is extraction of the aspect toward which sentiment is directed. As a result of message size limitations, Twitter users often share their opinion on events described in linked news stories that they find interesting. Therefore, a natural language analysis of the linked news stories may provide useful information that connects the Twitter-expressed sentiment to its aspect. Our goal is to monitor sentiment towards political actors by evaluating Twitter messages with linked event code data. We introduce a novel link-following approach to automate this process and correlate sentiment-bearing Twitter messages with aspect found in connected news articles. We compare multiple topic extraction approaches based on the information provided in the event codes, including the Goldstein scale, a simple decision tree model, and spin-glass graph clustering. We find that while Goldstein scale is uncorrelated with public sentiment, graph-based event coding schemes can effectively provide useful and nuanced information about the primary topics in a Twitter dataset.

1 Introduction

The use of publicly accessible Internet social networking sites (social media) now occupy a key role in the spread of information. Often this spread of information is paired with affective statements indicating sentiment [1]. General public sentiment within a region can change with events that effect its population. In particular, recent international events such as natural disasters, the cascade of protests and revolutions in the Arab world, and terror attacks have uncovered the utility of social networking sites for understanding social and political unrest. Effectively monitoring changes in sentiment toward political events may reveal trends in public opinion that indicate social and political instability.

A key component to understanding public opinion is extraction of the aspect toward which a given sentiment is directed. A metric to quantify the sentiment that social media users express towards political actors and/or politically relevant events would be useful to measure how much these actors or events influence affected populations. Twitter is a micro-blogging service that limits messages to 140 characters in length. Because of this limitation, the aspect of directed sentiment is often contextually embedded in Uniform Resource Locator (URL) links to other sources of information. For example, Twitter users commonly share an opinion on events described in linked news stories.

Open source automated political event coding schemes and datasets provide information about interactions between state and non-state actors in online news stories. The Open Event Data Alliance's Phoenix dataset [2] relies on the open source Python Engine for Text Resolution and Related Coding Hierarchy (PETRARCH) project, an automated coding library utilizing natural language processing (NLP), a lexical reference consisting of event/actor ontologies and verb/noun phrase dictionaries, and the Conflict and Mediation Event Observations (CAMEO) [3] coding scheme, to output source/target actors, an event that took place between actors, a location, and date of the event. These datasets have been studied to forecast political instability, identify trends such as escalation of conflict between actors, and monitor interaction between countries.

We develop an innovative link-following approach to automate the process of identifying sentiment and aspect in contextually ambiguous tweets by correlating sentiment-bearing Twitter messages with aspect found in shared news articles. We compare multiple topic identification approaches based on information provided in the event codes, including the Goldstein scale [4], a simple decision tree model [5], and spin-glass graph clustering [6]. We find that while Goldstein scale is uncorrelated with public sentiment, graph-based topic modeling does yield aspects directly tied to the tweets themselves, with information that cannot be gleaned from automated content analysis of the tweets alone. Correlating extracted topics with corresponding measures of public opinion within the tweets themselves may help identify trends for studying and predicting social and political changes.

2 Methodology

To develop an approach for correlating social media and political event data, we used a self-collected dataset from the Twitter streaming Application Programming Interface (API). We used twenty-five broad English hashtag-based search terms related to European governmental organizations and known events to collect Twitter data. Using broad, geopolitically relevant terms allows us to collect a large dataset likely containing a substantial number of URL links to politically relevant news stories. The search terms were chosen to facilitate manual grouping of tweets into one of four manually chosen topics: (1) EU Referendum in the UK (Brexit), (2) Migrant Crisis, (3) the North Atlantic Treaty Organization (NATO), and (4) Russia. We also allowed for an "Other" topic, encompassing anything not easily classified in the previous categories. 123,649 total tweets were

Table 1: Event distributions by ground truth topic

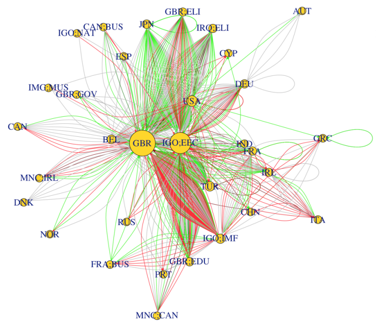
	Brexit	Migrant Crisis	NATO	Russia	Other
Tweets with Extracted Events	1,582	410	338	981	744
Unique Actor/Event Dyads	409	218	186	475	434
Unique Actor Dyads	166	132	85	202	267

collected from April 26, 2016 through May 23, 2016. To identify sentiment of the tweets, we employed the lexical approach introduced by Musto et al. (details found in [7]). As this approach relies on English, we used a logistic regression model across infinite-length character n-grams as language detection [8] to filter out non-English tweets, leaving 94,570 of the 123,649 collected tweets.

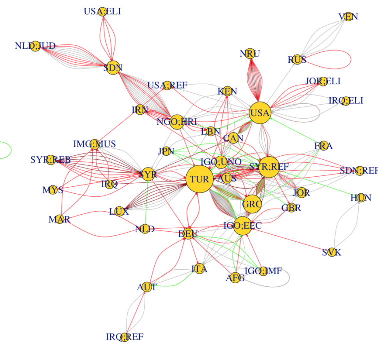
We then determined English tweets contained URL links, and followed these links to trusted news outlets with a web scraper. PETRARCH [2] was applied to the retrieved articles to extract geopolitical events with automated coding. From the linked URLs, we identified 4,323 unique events. Several tweets referenced the same news articles, yielding 10,287 English tweets with a corresponding event extracted from the linked URL. Of those, we found 4,055 were identified as subjective messages (a non-neutral sentiment score, heuristically $> |0.5|$). The event coding scheme yields actor dyads, CAMEO-based event type codes [3], and a corresponding Goldstein scale [4]. The Goldstein scale represents an ordinal measure ranging from extreme conflict (-10.0) to extreme cooperation (8.3). If sentiment were directly tied to the degree of conflict or cooperation, one could expect that negative sentiment corresponds to a high degree of conflict and positive sentiment corresponds to a high degree of cooperation.

The actor dyads represent the two primary state or non-state actors involved in the extracted event, each containing a country code with up to two role codes. We used the combined country code and secondary role code of the CAMEO actor dictionary to define distinct actors. The set of dyads can be represented by a directed graph, where the edge weight can be either sentiment or Goldstein scale. We utilized multiple reviewers to form a consensus manual, "ground truth" label for each tweet by one of the five topics (ignoring event codes but reviewing tweets and corresponding links). We then compared these labels to those generated by automated topic modeling approaches using the event graph dyads.

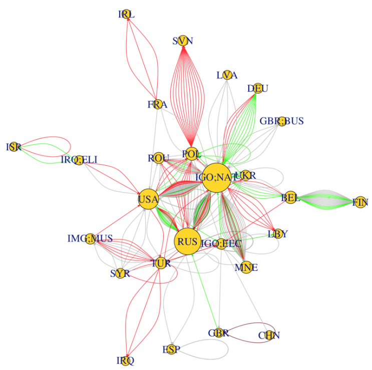
First, we employed a simple, supervised decision tree model [5]. We used five-fold cross-validation to perform training and testing, with individual actors and actor dyads as features, and a conditional inference based framework [9] for training. Alternatively, we applied an unsupervised graph clustering approach to identify individual topics based on the graph constructed from actor dyads. We employed the spin-glass graph clustering algorithm [6] based on the Potts model [10] of interacting spins on a crystalline lattice. Advantages to this graph clustering approach are that it naturally leads to known modularity measures [11] and allows for both directed and weighted networks.



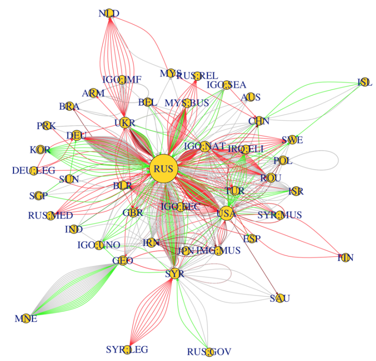
(a) Brexit



(b) Migrant Crisis



(c) NATO



(d) Russia

Fig. 1: Actor graphs by topic, with edges colored by five sentiment categories from very negative (dark red) to very positive (dark green)

3 Results

A breakdown of the 4,055 tweets by manually labeled topic are shown in Table 1. The *Brexit* and *Russia* topics make up a majority of the messages (39.0% and 24.2%, respectively). The large number of *Brexit* tweets is a result of the immediacy of the Referendum vote taking place in the United Kingdom and the English language restriction. The *Russia* topic was fairly broad and covered a wide variety of subtopics, including the interaction between NATO and Russia, which made manual labeling of the "ground truth" data slightly subjective. Tweets categorized as *Migrant Crisis* also covered a range of subtopics corresponding to events taking place in Syria, Greece, Sudan, Iraq, and Nauru, among others. Table 1 also shows the unique actor/event dyads and actor dyads (irrespective of event type) in the event coded news articles from extracted tweet links.

The Pearson product-moment correlation between sentiment and Goldstein scale was 0.094, not indicative of a strong correlation. Hence, the amount of conflict or cooperation of the events does not reflect the public opinion regarding

Table 2: Precision, recall, and f-measure metrics for topic modeling

(a) Decision tree classification

Class	Precision	Recall	f_1
Brexit	0.88	0.96	0.91
Migrant Crisis	0.75	0.81	0.78
NATO	0.72	0.56	0.63
Russia	0.81	0.84	0.83
Other	0.79	0.62	0.70
Macro-Avg Total	0.79	0.76	0.77

(b) Spin-glass graph clustering

Class	Precision	Recall	f_1
Brexit	0.92	0.94	0.93
Migrant Crisis	0.70	0.50	0.58
NATO / Russia	0.89	0.95	0.92
Macro-Avg Total	0.84	0.72	0.81

how positively or negatively they feel about the event. Someone may be pleased about conflict and angry about cooperation, or vice versa. The two measures provide distinct and useful information; however, one cannot be substituted for the other for analyses.

The resulting event subgraphs (not including the *Other* category) are displayed in Fig. 1. The CAMEO codes for geopolitical actors are the node labels, with node size reflecting the frequency that an actor appeared in the dataset. Edges represent individual tweets with links carrying a resultant event code and are colored by the sign of the sentiment. Several observations can be made from these graphs, including the centralization of *Brexit* (around the United Kingdom and European Union), decentralization of *Migrant Crisis* (composed of several subtopics), complexity of *Russia*, and the interdependence of *Russia* and *NATO*. Precision, recall, and f-measure for the supervised decision tree model are provided in Table 2a. The total accuracy for the model was 82.4%. The model is fairly accurate at identifying topics using the actor dyad information from the extracted event codes alone, without using any tweet content.

Spin-glass clustering resulted in ten total topics found with the maximum number of communities set at 20. Four of the generatively derived topics could be easily identified as representative of a *Brexit* cluster, a *Russia/NATO* combined cluster, and two separate clusters around *Migrant* issues. The remaining clusters could all be classified as part of the manually labeled *Other* category, with node sizes no greater than 11. As mentioned previously, the *Russia* and *NATO* labeled topics were interdependent, so it is not surprising that they were clustered together. Furthermore, the subtopics within the *Migrant Crisis* could easily be part of different topic areas if not forced into the same one through pre-determined classes. This unsupervised approach may actually produce topical breakdowns that are more indicative of the data, as the chosen topic areas were broad and with unrepresented complexities in the labels.

To evaluate the accuracy of the unsupervised algorithm, we distilled the topic classes into three categories (*Brexit*, *NATO/Russia*, and *Migrant Crisis*), as the former *NATO* and *Russia* topics are intertwined and the nuance of the individual clusters found by the spin-glass algorithm are lost in the manually labeled *Other* category. The total accuracy for the model was 88.7%. Precision, recall, and f-measure metrics are provided in Table 2b. The supervised and unsupervised event-topic modeling approaches yield somewhat similar prediction capabilities. The spin-glass method provides the added advantage of using the natural structure of the manifolds in the data itself rather than forced manifolds based on the broad manual labels.

4 Conclusion

The brevity of tweets makes identifying sentiment aspect in Twitter data difficult. However, many geopolitically-relevant tweets expressing opinion have links to news articles with more information about the topic and/or event(s). Event coding combined with topic modeling can be used to inform public sentiment aspect. This form of sentiment aspect may be used to identify the political actors involved as identified through event coding. An unsupervised graph clustering approach provides accurate results, with the added benefits of not requiring training data and an ability to capture nuance relationships between topics that may be missed by subjective manual labels. Both of these approaches were more successful than traditional topic modeling approaches such as latent Dirichlet allocation (LDA) [12], which was unable to achieve a macro-averaged f_1 score greater than 0.5 using the tweet content, the actor dyads, or the original linked news article content.

References

1. Ringsquandl, M., Petkovic, D.: Analyzing Political Sentiment on Twitter. AAAI Spring Symposium (2013)
2. Schrodt, P., Beieler, J., Idris, M.: Three's a Charm?: Open Event Data Coding with EL:DIABLO, PETRARCH, and the Open Event Data Alliance. International Studies Association Meeting (2014)
3. Schrodt, P. A. : CAMEO: Conflict and Mediation Event Observations Event and Actor Codebook. Pennsylvania State University (March 2012)
4. Goldstein, J.S.: A Conflict-Cooperation Scale for WEIS Events Data. J. of Conflict Resolution, 36(2) (1992)
5. Quinlan, J.R.: Induction of Decision Trees. Machine Learning, 1 (1986)
6. Reichardt, J., Bornholdt, S.: Statistical Mechanics of Community Detection. Phys. Rev. E, 74 (2006)
7. Musto, C., Semararo, G., Polignano, M.: A Comparison of Lexicon-based Approaches for Sentiment Analysis of Microblog Posts. 8th International Workshop on Information Filtering and Retrieval (2014)
8. Shuyo, N.: Short Text Language Detection with Infinity-Gram. NAIST Seminar (2012)
9. Hothorn, T., Hornik, K., Zeileis, A.: Unbiased Recursive Partitioning: A Conditional Inference Framework. J. of Comp. and Grap. Stat., 15(3): 651-674 (2006)
10. Wu, F.-Y.: The Potts Model. Reviews of Modern Physics, 54: 235 (1982)
11. Newman, M.E.J., Girvan, M. Community Structure in Social and Biological Networks. Proc. Nat. Acad. Sci., 99: 7821-7826 (2002)
12. Blei, D.M., Ng, A.Y., Jordan, M.I. Latent Dirichlet Allocation. J. Machine Learning Research, 3(4-5): 993-1022 (2003)