

Constellations on Social Representation of Candidates on the Landscape of Cooperation and Conflict during the 2016 National Election Campaign Period in the Philippines

By

Maria Regina Justina E. Estuar, Ph.D., John Noel C. Victorino, MSCS, , Kennedy Espina, May Gizelle Galarce, Rhene Muez, Jann Alfred Quinto, Joesalyn Relativo, Rey Rodriguez, Reuel Joab Yap, Marlene de Leon

Ateneo Social Computing Science Laboratory, Ateneo de Manila University

Abstract

On May 9, 2016, the citizens of the Philippines will exercise their power to choose the national leaders whom with high hopes that the chosen ones will uplift the lives of countrymen during the next 6 years. Like any typical election season, several issues and controversies abound such as legitimacy of source code review, electronic voting machines, hacking of the Commission on Election website, and data leaks. Moreover, social media has played a major role in the promotion (or not) of presidential and vice presidential candidates. Since the official filing of candidacy last October 12, 2015 until the official start of the campaign period beginning February 9, 2016 until the last day of the campaign on May 7, 2016, there have been reports of verbal and material conflict and cooperation related to the upcoming elections from online media sources as well as social media. Both media sources have also exposed candidates to criticism and praise creating spheres of social representation for each one. This study takes on the challenge of producing a quantifiable social representation using qualitative data through multidimensional scaling and geospatial mapping on conflict, cooperation, candidates and respective social representations during the 2016 National Election. The challenge is in fusing textual data coming from two sources: GDELT database and Twitter to discover if there are similar dimensions on reports of conflict and cooperation in both sources, with GDELT data considered official and considering twitter data as natural conversation.

Using the GDELT database, event location was set to Philippines, and data sets were generated using cooperation and conflict as event code, verbal and material as event class. A script was created to filter out all non-election related links produced by the GDELT database using names, nicknames, and common labels of candidates as keywords. A total of 17,581 records on material conflict, 12764 records on verbal conflict, 12550 records on material cooperation, and 100332 records on verbal cooperation were obtained from the GDELT database. Top themes for material conflict during elections include: *Use conventional military force, Arrest, detain, or charge with legal action* and *Fight with small arms and light weapons*. Top theme for verbal conflict was *accuse*. In material cooperation, the top theme was *investigate*

while *make a visit, praise or endorse, host a visit* was the top theme for verbal cooperation.

Aside from using base data from the initial GDELT search, a second level of search was performed to obtain more meaning from the content of the source websites. A script was created to crawl the web to obtain textual information from the web sources indicated in the GDELT output producing a total of 2800 records on material conflict, 3692 records on verbal conflict, 1833 records on material cooperation, and 20347 records on verbal cooperation. Tweets pertaining to the national elections and candidates were also harvested from the start of the campaign period until April 25, 2016.

The final dataset for analysis consisted of documents from GDELT and documents from the Twitter dataset. Both datasets were subjected to standard data cleaning. In addition, to retain only adjectives, an additional step of filtering was made so that the final wordlist contains only candidates and adjectives. Word association was used to generate the initial candidate – adjective scores, from which the final candidate-adjective matrix was produced. Multidimensional scaling was used to produce the constellations of candidates and adjectives on verbal and material categories within the conflict and cooperation landscape. Areas of material and verbal conflict and cooperation were visualized as it relates to the social representation of each candidate. A model on predicting the event code of the tweets was developed using GDELT dataset as gold standard.