SOCIAL MEDIA, U.S. PRESIDENTIAL CAMPAIGNS, AND PUBLIC OPINION POLLS: DISENTANGLING EFFECTS

Jennifer Stromer-Galley
Syracuse University

Patricia G. C. Rossini
Federal University of Minas Gerais, Brazil

Jeff Hemsley
Syracuse University

Kate Kenski
University of Arizona

Feifei Zhang
Syracuse University

Lauren Bryant
University at Albany

Bryan Semaan
Syracuse University

Twitter and Facebook have now "come of age" for strategic communication by presidential campaigns in the United States. For example, in 2012, the Obama campaign estimated they could reach 95% of the voting public through Facebook (Obama Legacy Report, n.d.). While neither Twitter nor Facebook have increased their user bases substantially since 2012, they are an important source of information for the electorate during an election, including this election cycle (Gottfried, Barthel, Shearer & Mitchell, 2016).

Because social media is now an important site of communication for campaigns and the electorate, we seek to understand the types and nature of messaging that campaigns produce. Of specific interest is the relationship between messaging on social media and public opinion polls. Prior research suggests that strategic messages in television ads change based on candidate standing in the polls (Jamieson, 1992). As races tighten, they

tend to become more negative (Buell & Sigelman, 2008; Hassell & Oeltjenbruns, 2015; Lau & Pomper, 2004). Two recent studies suggest that campaigns mirror what they do offline and online (Druckman, Kifer, & Parkin, 2009; Krupnikov & Easter, 2013). This leads us to predict that presidential candidate message strategies will change based on their standing in vote-intention polls, which ask the public who they would vote for if the election were held today.

Methods

We collected Facebook and Twitter messages of all 17 Republican and 7 Democratic primary. The period of analysis starts from when the candidates announced they were running early in 2015 until they dropped out, or April 1st, 2016, whichever came first. The analysis focuses on the surfacing and primary stages of the campaigns because of the large number of candidates running, which allows for the highest number of data points to examine the relationship between strategic messaging and polling. We use national public opinion polling data from RealClearPolitics.com, which provides a rolling average of an aggregate of national polls.

We developed a codebook to categorize tweets through deductive and inductive analysis of Twitter messages by the candidates (Krippendorff, 2003). In this study, we focus on strategic messages. These are operationalized as any message that is about the candidate or their opponent that focuses on policy and issues, or character and personality, including policy messages around voting, and standing in the public opinion polls. We defined two sub-categories under strategic messages: attack and advocacy messaging. An attack message is operationalized as a message that criticizes the opponent or opposing administration or party on their personality, leadership skills, past behaviors, family, policy issues, campaign events, or any other negative focus on the opponent, or their campaign, surrogates, or family. An advocacy message is operationalized as a message that advocates for the candidate, highlighting their strengths as a leader, describing their prior policies or personal history, describing or featuring their family, describing or highlighting their current and future policy positions, or featuring their positive personality characteristics, is an advocacy message.

We will use machine learning to categorize the entire corpus of Facebook messages and tweets. Human coders have annotated strategic messages from both Facebook and Twitter using 2014 Gubernatorial campaign data. Once annotators reached at least 75% agreement on all categories, they reconciled differences to generate gold-labeled data. To date, our annotators have annotated and adjudicated 1,438 strategic tweets, with 958 advocacy tweets and 480 attack tweets; and 856 Facebook strategic messages, with 537 advocacy messages and 319 attack messages. We use these gold standard data as training data to build algorithms for advocacy and attack message detection.

For algorithm building, we performed a number of experiments in Scikit-learn, a python-based, command-line machine learning package. All classification tasks were evaluated with 10-fold cross validation. For strategic messages in Twitter (see Table 1), the best micro-averaged F value is 0.77, by using SVM classifier with features, including Boolean, @_username, and numbers of @_username included in one tweet. For

Facebook data (see Table 2), the best micro-averaged F value is 0.83, by using a SVM classifier with a Boolean feature. By comparison, the majority baseline for Twitter strategic sub-category is 0.67, and 0.63 for Facebook. All the micro-averaged F values reported are much higher than baseline scores. This suggests that the machine-coding algorithms have been trained to predict these categories well. As a further step, we will test the reliability of our current best models and then apply them to predict strategic messages in the presidential campaign, with an additional comparison with human annotation to ensure valid and reliably tagged data.

We will use time-series analysis to examine the relationships between strategic message types and standing in public opinion polls.

Findings and Implications

Because the primary period commenced in late February, 2016, and has not concluded yet, we are still collecting data. As such, we do not have an indication of the results. The results of this research may provide evidence on the relationship, if there is one, between strategic messaging on social media and candidate standing in the polls. Understanding the dynamic between polls and strategic messaging would contribute to theories of digital campaigning and political communication, and may support prior research that suggests that presidential campaign messaging is dynamic and is shaped by the larger media environment (Denton, 1998). Finally, the results, may provide caution to researchers who aggregate campaign social media messages using big data analytics into categories without factoring in the role of time and the external factors, such as polling data that shape strategic messaging in social media.

References

Buell, E. H., & Sigelman, L. (2008). *Attack politics: Negativity in presidential campaigns since 1960*. Univ Pr of Kansas.

Denton, R.E. (1998). Communication variables and dynamics of the 1996 presidential campaign. In R. E. Denton (Eds.), *The 1996 presidential campaign: A communication perspective* (pp. 1-50). Westport, CT: Praeger.

Druckman, J. N., Kifer, M. J., & Parkin, M. (2009). Campaign communications in US congressional elections. *American Political Science Review*, *103*(03), 343–366.

Gottfried, J., Barthel, M., Shearer, E., & Mitchell, A. (2016, Feb. 4). The 2016 presidential campaign – a news event that's hard to miss. Pew Research Center. Retrieved, March 1, 2016 from http://www.journalism.org/2016/02/04/the-2016-presidential-campaign-a-news-event-thats-hard-to-miss/.

Hassell, H. J., & Oeltjenbruns, K. R. (n.d.). When to Attack: The Trajectory of Congressional Campaign Negativity.

Jamieson, K. H. (1992). Dirty politics. *Deception, Distraction, and Democracy. New York*.

Krippendorff, K. H. (2003). *Content analysis: An introduction to its methodology* (2nd. Ed.). Thousand Oaks, CA: Sage.

Krupnikov, Y., & Easter, B. C. (2013). Negative Campaigns. *New Directions in Media and Politics*, 100.

Lau, R. R., & Pomper, G. M. (2004). *Negative campaigning: An analysis of US Senate elections*. Rowman & Littlefield.

Obama Legacy Report. (n.d.). Retrieved February 28, 2013 from http://secure.assets.bostatic.com/frontend/projects/legacy/legacy-report.pdf.

Tables

Strategic Messages Sub-category	Precision	Recall	F1	NO. of messages
Advocacy	0.85	0.79	0.82	958
Attack	0.63	0.72	0.67	480
Micro-average	0.77	0.76	**0.77**	1438

Table 1: Machine prediction performance for Strategic Messages in Tweeter

Strategic Messages Sub-category	Precision	Recall	F1	NO. of messages
Advocacy	0.85	0.89	0.87	537
Attack	0.80	0.73	0.77	319
Micro-average	0.83	0.83	**0.83**	856

Table 2: Machine prediction performance for Strategic Messages in Facebook