Noah Segal-Gould

Prof. Maria Cecire

Introduction to Media

8 December 2016

2016 Presidential Election Candidate Tweets Classification and Topic Modeling

Over the course of the semester, culminating in the successful presidential campaign of

Donald Trump in November, I struggled to make sense of the media consumption and

production which interchangeably pervade my life. I often feel anesthetized. The sheer quantity

of content I consume and produce makes no singular moment, idea, medium, or piece of media

necessarily seem influential on anything I say or do. Why, then, did I feel like I had played a role

in what I perceived to be a failed election? When I looked to social media before the night of

November 8th, I sought political humor. After, I began to regret what I had done. More than a

racist, more than a misogynist, and more than a homophobe, across the spectrum of admittedly

liberal media I consumed, Donald Trump was a joke. Social media left me underprepared to

consider his presidency as even a potential outcome. It upset me that I had not considered it

precisely because I told myself I was an informed voter. Evidently, I was only informed enough

to consider outcomes I desired. The disappointment in myself and the liberal political and social

media I consumed, which then followed, shaped a question: What differentiates political

discussions on social media? I wanted to know what role I played in them, but I especially

wanted to know how in my efforts to give and take from these services, I never conceived of

boundaries between political parties, candidates, and voters quite the way I would have in

physical political discourse.

Who tends to say what? From a computational perspective, I broke this question up into two parts. The original 'who' is specified by five major candidates for president of the United States: Hillary Clinton, Donald Trump, Bernie Sanders, Gary Johnson, and Jill Stein. The outlier among these five who was not even running for president on November 8th is Bernie Sanders, who conceded the democratic nomination to Hillary Clinton. I chose to keep him simultaneously for his involvement in the election prior to his concession and his consistently massive social media audience after it. For each candidate, I found the Tweet in which they announced their candidacy, and then accessed the text of each following Tweet until the end of the day on November 8th. These five datasets were combined and categorized by the author of the Tweet. In application of text classification algorithms[1], I built a predictive and probabilistic model to provide distributions of likelihood on the classification of user input into each of the five categories. I went on to apply topic modeling[2] to extract the most relevant content from each dataset, and then visualize the distribution of the topics represented in user input. When I visit intro-to-media.n04h5g.me, leave the default text in the user input box, and click "Submit as Naive Bayes, the result is read as such: 28.26% like Bernie Sanders, 24.97% like Donald Trump, 24.36% like Hillary Clinton, 19.34% like Jill Stein, and 3.07% like Gary Johnson. Then, the bar chart shows the most prevalent topic represented in the input interpreted as a Sanders Tweet is topic #3. For Sanders, that topic in particular contains words like "change, climate, vote, energy,

[1] "Sklearn.naive_bayes.MultinomialNB." Sklearn.naive_bayes.MultinomialNB — Scikit-learn 0.18.1 Documentation. Accessed December 09, 2016. http://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html.

[2] "Sklearn.decomposition.LatentDirichletAllocation." Sklearn.decomposition.LatentDirichletAllocation — Scikit-learn 0.18.1 Documentation. Accessed December 09, 2016. http://scikit-learn.org/stable/modules/generated/sklearn.decomposition.LatentDirichletAllocation.html.

and trade." The text classification model provides the 'who' and the topic model provides the

'what' in "Who tends to say what?"

This web application is an experiment and exploration and should not be taken to

represent fact. It feels very misleading for it to represent any textual idea as being a distribution

of presidential candidates, precisely because the type of content users are drawn to input rarely

resemble Tweets actually made by the them. For this reason, I added the option to either submit

known hashtags from their campaigns in the dropdown menu, or a randomly selected Tweet

which was actually shared by a candidate. This program is not objective and the models which

lie under its hood are flawed and dumbed down to work quickly for fast demonstration. The

models themselves are not nearly as interesting to me as the response this project garnered. Many

individuals to whom I showed the site chose to input content that would not be very polarizing

among candidates but instead chose to give it sentences which would never have been Tweeted

by a candidate. In that way, the site works a lot like participatory performance art. This reminded

me of a very influential passage from Walter Benjamin's *The Work of Art in the Age of*

Mechanical Reproduction[3]:

> Mankind, which in Homer's time was an object of contemplation for the Olympian gods, now is one for itself. Its self-alienation has reached such a degree that it can experience its own destruction as an aesthetic pleasure of the first order. This is the situation of politics which Fascism is rendering aesthetic. Communism responds by politicizing art (114).

Taken as politicized art which seeks to respond to the fascist ideals president-elect Trump and

his chosen accomplices have perpetuated, perhaps this project is reactionary. I do not know what

Benjamin would make of this kind of art. It strikes me as transformative on the aura of its

---

[3] Ouellette, Laurie. The Media Studies Reader. London: Routledge, 2010.

constituents. It recontextualizes all the tens of thousands of Tweets which went into it, and shows something different entirely.

The decisions users make on the input can then be perceived in the output. Surely there exist topics which Hillary Clinton never took to Twitter to discuss. Those are the topics which the program necessarily weight more highly in the distribution precisely for lack of usage. Users do not know this, and thus the nonsense they input amuses them as they perceive it as fact. Every time a user asks the site about a body of text, a Twitter bot in the background Tweets whatever they thought to resemble a Tweet from a presidential candidate along with the pie chart that visualizes the probabilistic distribution of candidates represented in that Tweet via @intro_to_media. Is this a dialogue in which users even have a right to participate? When *Convergence Culture*[4] was published in 2008, Henry Jenkins sought to address that concern:

> On all sides and at every level, the term participation has emerged as a governing concept, albeit one surrounded by conflicting expectations. Corporations imagine participation as something they can start and stop, channel and reroute, commodify and market. The prohibitionists are trying to shut down unauthorized participation; the collaborationists are trying to win grassroots creators over to their side. Consumers, on the other side, are asserting a right to participate in the culture, on their own terms, when and where they wish. This empowered consumer faces a series of struggles to persevere and broaden this perceived right to participate (175).

My project recontextualizes the politics of the election with very clear social media borders, and forces users to participate in situating themselves within those borders as consumers and producers of political social media. Writing *Harry Potter* fanfiction and deciding who to vote into office as president of the United States may not appear similar, but the participants who take part in the resultant media are similarly challenged. Heather Lawver subverted systems which

---

[4] Jenkins, Henry. Convergence Culture: Where Old and New Media Collide. New York: New York University Press, 2006.

would prefer she not "participate in the culture," but even while these politicians necessarily functioned by encouraging and often demanding social media participation, the borders between them on social media that they chose to propagate surely sought to do the same as Warner Bros. sought to do to Heather's *The Daily Prophet*: control precisely which individuals possess authority over the messages contained within a piece of media. This project turns those borders and that control into something nonhuman. The program does precisely what real Twitter users were taught not to do: it easily manages to traverse these boundaries that have been set up to restrain users, and seeks to remind them that they too may choose to participate as they see fit.

Bibliography:

"Sklearn.naive_bayes.MultinomialNB." Sklearn.naive_bayes.MultinomialNB — Scikit-learn 0.18.1 Documentation. Accessed December 09, 2016. http://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html.

"Sklearn.decomposition.LatentDirichletAllocation." Sklearn.decomposition.LatentDirichletAllocation — Scikit-learn 0.18.1 Documentation. Accessed December 09, 2016. http://scikit-learn.org/stable/modules/generated/sklearn.decomposition.LatentDirichletAllocation .html.

Ouellette, Laurie. The Media Studies Reader. London: Routledge, 2010.

Jenkins, Henry. Convergence Culture: Where Old and New Media Collide. New York: New York University Press, 2006.