Matthew Daniel and Cortlandt Bursey-Reece

CS 221

Introduction

Americans are currently experiencing what seems to be an unprecedented level of ideological division, with high levels of political polarization occurring on both sides of the aisle. This division may be more immediately apparent on social media than anywhere else, where ideological echo chambers can easily fan the flames of a user's beliefs. Social media has the power to show people content that they find politically agreeable: the goal of most social media sites is to get people to remain on the site longer, but a side effect is the creation of an ideological feedback loop that reinforces their beliefs. This feedback loop, as we saw in the 2016 election, presents an opportunity for exploitation.

In the election, Russian bot and troll accounts on Twitter contributed to this phenomenon by flooding the website with politically charged and often incorrect information. The problem was so widespread that an estimated 20% of tweets about the election were fake. In 2017, a study found an estimated 48 million fake accounts (1) on the site. Twitter has taken steps to remove these accounts and limit the spread of disinformation, but new fake accounts are continually being created, and this has increased in intensity in the period leading up to the 2018 US midterms. The problem has also spread to other countries, and fake accounts have influenced elections in France, Brazil, and many other countries around the world.

While many of these fake accounts are directly supportive of the causes they intend to help, others take a more insidious approach, posing as the opposition in an attempt to discredit or divide the other side. This makes easy detection of these accounts difficult, but with social media political engagement at a high, ensuring that social media users have access to feeds free of falsified information is more crucial to the democratic process than ever before.

Data

We used two datasets when collecting preliminary data: a russian-troll dataset with 200,000+ tweets, and a Democrat-/Republican-congressmember with 90,000 + tweets. In our algorithm, we used a unigram weights vector with select words to try and classify tweets as 1) troll or not, and 2) left-leaning or right-leaning. For each tweet, we looked at each word and added the weight to the tweet's score. We labeled it as +1 if the score was positive, -1 if the score was negative, and no_class if the score was 0. When classifying trolls, a +1 indicates troll and -1 indicates not a troll, and when classifying political leaning, a +1 indicates right-leaning and -1 indicate left-leaning. Our weight vectors were hand selected based on what we believed would indicate a left/right and troll/not troll split. Here's sample of the political bias vector:
{'rigged': 1, 'clinton': 1, 'liberals': 1, 'benghazi': 1, conservatives': -1, 'trump': -1, 'notmypresident': -1, 'hope': -1, }.
And here's a sample of the troll-classification vector:

```
{'nigga': 1, 'nastywoman': 1, 'maga': 1, 'racist': 1, 'fuck': 1, 'civic': -1, 'vote': -1, 'hope': -1, 'respect': -1, 'change': -1}
```

The classifications with "Democrat/Republican" as the first string came from the congressmember dataset.

Sample of Political Classification:

-1 Democrat i'm proud to voice my support for small businesses as we celebrate their contributions to our neighborhoods. today,… https://t.co/zbkuin

+1 Republican i support pres. @realdonaldtrump's #rescission package, and would like to see where the democrats stand on billions… https://t.co/uoqogz

+1 rt @shareblue: pence and his lawyers decided which of his official emails the public could see

-1 rt @dianerainie: hey @hillaryclinton this message is for you. pack it up & go home hillary

no_class the most honest hillary's poster i've seen 😂

Sample of Troll Classification:

no_class Republican today we honored all the heroic men &amp; women of law enforcement who lost their lives on the line of duty at the nati… https://t.co/iejjzlmjdg

+1 Democrat rt @naacp: pardon of arpaio is explicit embrace of the racist policing practices that leave communities fearful of very ppl who should prot…

-1 Democrat i'm deeply disappointed to see this republican-controlled congress vote to ship radioactive nuclear waste across th… https://t.co/wucwmux

no_class one of the ways to remind that #blacklivesmatter #blackpressday

+1 rt @wokieleaks: congratulations to donald trump, our first president to be metal as fuck

-1 rt @sunkissdnerd: is this the leader that will fight isis &amp; take our country out of the ditch obama put us in? #votetrump #imwithyou https:…

Oracle, Challenges, and Gaps

Our simple approach classified 382 Republicans right, 1896 Republicans wrong, and 42K not classified, while 4155 Democrats were classified right, 284 wrong, and 37K not classified. Meanwhile, 1500 trolls were labeled properly, 5000 wrong, and 19K not classified. The major challenge is that many of our tweets (especially in troll classification) are unclassified or classified wrong, and Republicans were more likely to be mislabeled. There is also isn't much data that finds trolls and classifies them as left- or right-wing. We hope our project will close that gap.

Future Methods

For the future, we plan on using naive bayes to 1) determine whether or not a tweet/user is a troll, and 2) if they are, determine if the troll is a right-wing troll (either posing as a left-wing or right-wing nut) or a left-wing troll (either posing as a left-wing or right-wing nut). Past 221 projects (2, 3) have used naive bayes for political classification, so we are confident the approach will properly classify left- and right-wing tweets. We will look at known left- and right- wing astroturf tweets to determine if the classified tweets are genuine left- and right-wing. We also believe naive bayes will work well in grouping trolls and non-trolls together. We hope our project will allow people to more easily spot trolls and more readily classify them as left- or right-wing.

Matthew Daniel and Cortlandt Bursey-Reece

CS 221

References

1. https://www.cnbc.com/2017/03/10/nearly-48-million-twitter-accounts-could-be-bots-says-study.html
2. http://web.stanford.edu/class/cs221/2018/restricted/posters/apham7/poster.pdf
3. http://web.stanford.edu/class/cs221/2018/restricted/posters/erjones/poster.pdf