CSCI 77800 - Ethics and Computer Science
November 30, 2022
Taylor Grant-Knight and Wei (Amanda) Lee

Fake News

"Fake news" has become a dominant term in our political and cultural discussions. Popularized by then candidate for president Donald Trump, this term is commonly used to discredit unfavorable media. During the 2016 election, and every election since, active social media users have seen ads, articles, and reposts of political content of dubious truthfulness. The question many began asking is: What portion of these articles posted online are actually false? In 2016, right before that election, BuzzFeed News produced a piece identifying statistics on articles published by major left, right, and center organizations. We have used this data to create a model that simulates the proportion of true, false, and mixed articles before and after reposting is taken into consideration.

The BuzzFeed News article identifies the top three hyper-partisan left, mainstream, and hyper-partisan right news organizations. For simplicity, we have lumped these nine organizations into three categories: left, center, and right. Over the course of a week, Buzzfeed News gathered a total of 2,282 article posts from these organizations. Quoting the results, "There were 1,145 posts from mainstream pages, 666 from hyperpartisan right-wing pages, and 471 from hyperpartisan left-wing pages." Each post was assigned a truth rating: Mostly false, mixed false and true, mostly true, and non-factual (like a meme). Notably, 0.0% of mainstream articles were deemed mostly false, while 4.7% of left-wing articles were false and 12.3% of right-wing articles were false.

To create our model, we defined each post to be a tuple of `(category, truth)`. The statistics from the article are turned into fraction values and compared to two random values. Based on these random values, a post is assigned a category (-1 for left, 0 for mainstream, 1 for right) and a truth rating (-1 for false, 0 for mixed or non-factual, 1 for true).

The article next introduces the observed resharing rate of articles. Each of the nine tracked organizations had vastly different levels of engagement. For simplicity, we took the sum of each of the median shares per post of each of the nine and consolidated this into three amounts of posts: 96 total for mainstream, 1,305 for right-wing, and a staggering 15,436 for left-wing. The orders of magnitude difference was a concern, and follow-up work on this topic should try to get other large samples of data to work with to confirm the findings of BuzzFeed News. However, this is still not unreasonable given the demographics of Internet users and led to interesting results for our model.

To model the resharing of articles, we developed a method that would take a post variable (i.e., a tuple of category and truth, both integers ranging from -1 to 1) and would return a list of copies of that post equal in size to the number of reshares that category of post would usually receive. Note that no direct data in the BuzzFeed News article was recorded for on reshares based on truth level, which would have added more to our model. The resulting reposts are then combined into one master list of all posts and the process would continue until each post has had its resharing modeled and recorded.

The most interesting conclusion from our model is that, because mainstream articles have such low engagement, the percentage of “mostly true” articles available online decreases before and after reposting is considered. That means that more average engagement and resharing online, the lower the proportion of true articles available in aggregate. See the chart on the next page for the results of three rounds running the model. Each round was generated using 100 initial posts (more than 100 began to slow the program considerably).

Our results serve to illustrate the implications of the BuzzFeed News article: Given the current nature of news media, the more engagement with articles online, the smaller the amount of reshared and thus total available true articles. Future research would benefit from breaking down the subgroups further and to take a closer look at the nature of engagement. Because of how Google, Facebook, and other search and social media algorithms work, it would also be beneficial to analyze the structure of a person's given "bubble" and how that affects the spread of news.

Initial: Left: 19 (19.0%) Center: 51 (51.0%) Right: 30 (30.0%) False: 8 (8.0%) Mixed: 26 (26.0%) True: 66 (66.0%) Ending: Left: 293303 (86.9%) Center: 4947 (1.5%) Right: 39180 (11.6%) False: 38710 (11.5%) Mixed: 142168 (42.1%) True: 156552 (46.4%)	Initial: Left: 20 (20.0%) Center: 37 (37.0%) Right: 43 (43.0%) False: 6 (6.0%) Mixed: 38 (38.0%) True: 56 (56.0%) Ending: Left: 308740 (83.8%) Center: 3589 (1.0%) Right: 56158 (15.2%) False: 21967 (6.0%) Mixed: 205069 (55.7%) True: 141451 (38.4%)	Initial: Left: 18 (18.0%) Center: 52 (52.0%) Right: 30 (30.0%) False: 3 (3.0%) Mixed: 22 (22.0%) True: 75 (75.0%) Ending: Left: 277866 (86.3%) Center: 5044 (1.6%) Right: 39180 (12.2%) False: 18049 (5.6%) Mixed: 111100 (34.5%) True: 192941 (59.9%)

Table: Three results showing the truth content of articles before and after reshares.

Reference:

https://www.buzzfeednews.com/article/craigsilverman/partisan-fb-pages-analysis