How fake news go viral?

Or why Bernie Sanders could replace President Trump with little-known loophole

Victor Kristof, William Trouleau

November 28, 2016

Abstract

The clic-bait headline of the title, published on November 14, 2016 by the Huffington Post, is clearly fake (and on purpose). It cleverly shades light on an issue that became mainstream with the recent US election: fake news propaganda on social networks. Indeed, an analysis of user engagement performed by BuzzFeed concluded that fake election news stories raised more engagement on Facebook than the top election story from 19 major news outlets combined [1]. The idea of this project is to analyze news articles from famous Facebook pages publishing articles related to the US elections and to measure the impact of misinformation on the elections.

Project goals

1 Data acquisition

The first goal of the project is to collect data from famous Facebook pages publishing articles related to the US elections. According to Buzzfeed [2], the following pages were really active:

- Mainstream pages: Politico, CNN and abc News Politics.
- Pro-democrate pages: The Other 98%, Addicting Info and Occupy Democrats
- Pro-republican pages: Eagle Rising, Right Wing News and Freedom Daily

Each collected article should include the main text messages in addition to user engagement metrics (i.e. the number of shares, comments, and reactions).

To write their article, BuzzFeed already collected this data for one week in September 2016. Then they manually annotated it to add a rating on the content truthfulness [3]. In addition to collecting the same data once again, our goal is to collect newer and/or older not annotated articles from the same sources to analyze them.

2 Data exploration

Once the data is collected, a first task should be to compare the data we collected ourselves with the data collected by BuzzFeed to certify their consistency. This first task is crucial to make sure that their findings can be trusted and are not propaganda themselves. However, given the level of transparency they provided on [2], we can think that their annotations are legitimate.

3 Data exploitation

The article published by BuzzFeed stopped at the data exploration part and did not investigate any machine learning method. The next goal of this project would then be to build a propaganda detector to automatically recognize fake news from genuine ones.

This problem is clearly challenging and we do not have a clear method to tackle the problem yet. However, we could think of the following pipeline:

- **Data preprocessing:** the first step would be to pre-process the data by converting the raw text messages into bag-of-word vectors. Other side information features like the existence of a link, photo or video could be considered.
- **Modeling:** We could then apply state-of-the-art methods of supervised classification to detect the truthfulness rating of unlabeled data.

4 Evaluation

The final task of the project is to evaluate the performance our model(s) to detect the truthfulness of articles. Assuming that the classes won't necessary be well balanced, the accuracy might not be a good metric. We may prefer to measure the F1-score to take into account both recall and precision. Additional metrics may be evaluated.

References

[1] BuzzFeedNews. This Analysis Shows How Fake Election News Stories Outperformed Real News On Facebook.

[2] BuzzFeedNews. Hyperpartisan Facebook Pages Are Publishing False And Misleading Information At An Alarming Rate.

[3] BuzzFeedNews Github repository. Fact-Checking Facebook Politics Pages `https://github.com/BuzzFeedNews/2016-10-facebook-fact-check`