Heinz 95-845: Project Proposal
Predicting False News on Twitter

Nicolas Banholzer NBANHOLZ@ANDREW.CMU.EDU
Heinz College
Carnegie Mellon University
Pittsburgh, PA, United States

Jasper Meijering JMEIJERI@ANDREW.CMU.EDU
Heinz College
Carnegie Mellon University
Pittsburgh, PA, United States

1. Proposition and Research Hypothesis

We have the dataset from Vosoughi et al. (2018). They analyze the spread of false and true news on Twitter and find that false news spread significantly farther, deeper and more broadly than true news. However, their analysis is mainly descriptive and they do not attempt to predict false or true news. Our aim is to use their data and try to predict whether news are false or true. Our precise research hypothesis is that *depth, range and virality are good predictors of the veracity of news.*

2. Relevance

Within social media, there are three main areas where credibility assessment is important: "(1) the detection of opinion spam in review sites, (2) the detection of fake news and spam in microblogging, and (3) the credibility assessment of online health information" (Viviani and Pasi, 2017). Regarding (2), not least since the 2016 US presidential, many are worried that social media users are especially subject to false news which can impact voting decisions and election results (Allcott and Gentzkow, 2017). To counteract, Lazer et al. (2018) propose two ways. First, empowering individuals in critical evaluation of content. Second, platform-based automated evaluation of the quality of content through algorithms, followed by false news detection and possibly prevention. While the first proposition probably requires long-term improvements in the educational system, the second proposition might be in closer reach, particularly when considering the data mining capabilities of Google, Facebook, Twitter and the like to predict user-relevant content (Lazer et al., 2018). Also, primarily social media platforms have strong incentives to check the veracity of content as they are especially susceptible to false news because they allow easy and cheap distribution of news (Shu et al., 2017) and because internal information on social media diffuses much faster than information coming from external sources (Yoo et al., 2016).

3. Contribution

Shu et al. (2017) give an overview of features and data mining techniques that can help analyze news and predicting their veracity. They distinguish between news content and

social context features. News content features describe the linguistic or visual style of the content and can for example inform on how false news are written. Indeed, Vosoughi et al. (2018) show for a smaller sample of their data that false news exhibit greater novelty. However, the main focus of their analysis is on how false news diffuse compared to true news. Since their analysis is mainly descriptive, we build on their work and propose a model to predict the veracity of news based on features measuring information diffusion and controlling for features of social context.

4. Data

The data by Vosoughi et al. (2018) consists of 126,000 stories in 4.5 million tweets by 3 million people between 2006 and 2017. The veracity of the stories is independently checked by 6 fact-checking organizations. Each story can have multiple "cascades", i.e. instances of the same story tweeted by independent users, and each cascade can have varying size, i.e. an instance of a story re-tweeted multiple times. The tweets comprise multiple categories, e.g. politics, business, entertainment, etc. We consider the following features:

(Y) binary outcome: news is true or false

(X) information diffusion: depth, size, breadth, virality[1]

(Z) social context: # followers, # followees, verification status, account age, engagement[2]

(W) population: active social media (Twitter) users between 2006 and 2017, only tweets in English language considered

5. Evaluation

Binary classification matrix ($\rightarrow$ accuracy, precision, ...), ROC, Cross-validation

6. Study Design and Methodology.

1. Pre-processing: Information diffusion features have to be computed.[3] the outcome variable is not binary and has to be recoded for binary classification.[4]
2. Sensitivity: It might make sense to predict not based on the full size of a cascade but rather set a cut-off, e.g. prune after 10 retweets.
3. Design: Split data in training (model fit), validation (tuning) and test set (evaluation).
4. Model: Neural networks (as they are among the most powerful ML models).
5. Robustness: The results could be compared by rumor category.

1. We could add further features about information diffusion or compute them differently than Vosoughi et al. (2018) (e.g. using different virality metric).
2. We refer to these as the control variables. In addition, we could also consider the time of the tweet.
3. Vosoughi et al. (2018) provide a python script to compute their features about information diffusion, but we probably have to compute them differently regarding our analysis.
4. It is coded as false, mostly false, mixed, mostly true, true. One way to create a binary outcome could be to exclude mixed stories and code mostly false as false and mostly true as true.

7. Limitations

- Representativeness: Social media pop. is not representative of whole pop. both w.r.t. demographics (PEW Research Center, 2018) and personality Correa et al. (2010).
- Omitted variable bias: The provided dataset does not include the raw text of the tweet, so adding content-related features is not possible using this dataset.
- Content type: Only written text or extracted from images considered, no audio/video.

References

Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–36, May 2017. doi: 10.1257/jep.31.2.211. URL http://www.aeaweb.org/articles?id=10.1257/jep.31.2.211.

Teresa Correa, Amber Willard Hinsley, and Homero Gil de Zúñiga. Who interacts on the web?: The intersection of users' personality and social media use. *Computers in Human Behavior*, 26(2):247 – 253, 2010. ISSN 0747-5632. doi: https://doi.org/10.1016/j.chb.2009.09.003. URL http://www.sciencedirect.com/science/article/pii/S0747563209001472.

David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. The science of fake news. *Science*, 359(6380):1094–1096, 2018. ISSN 0036-8075. doi: 10.1126/science.aao2998. URL http://science.sciencemag.org/content/359/6380/1094.

PEW Research Center. Social media fact sheet, February 2018. URL http://www.pewinternet.org/fact-sheet/social-media. Fact sheet is based on a survey conducted from January 3-10, 2018.

Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective. *SIGKDD Explor. Newsl.*, 19(1):22–36, September 2017. ISSN 1931-0145. doi: 10.1145/3137597.3137600. URL http://doi.acm.org/10.1145/3137597.3137600.

Marco Viviani and Gabriella Pasi. Credibility in social media: Opinions, news, and health information – a survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(5):e1209, 2017. doi: 10.1002/widm.1209. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1209.

Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, 2018. ISSN 0036-8075. doi: 10.1126/science.aap9559. URL http://science.sciencemag.org/content/359/6380/1146.

Eunae Yoo, William Rand, Mahyar Eftekhar, and Elliot Rabinovich. Evaluating information diffusion speed and its determinants in social media networks during humanitarian crises. *Journal of Operations Management*, 45:123 – 133, 2016. ISSN 0272-6963.

doi: https://doi.org/10.1016/j.jom.2016.05.007. URL http://www.sciencedirect.com/science/article/pii/S0272696316300419. Special Issue on Humanitarian Operations Management.