2017-18 Washington Research Foundation Fellowship Automatic Fake News Detection Research Plan

Deric Pang
Paul G. Allen School of Computer Science & Engineering
University of Washington
Seattle, WA
`dericp@cs.washington.edu`

1 Introduction

The impact of false news stories—popularly dubbed "fake news"—has been a specific concern following the 2016 U.S. presidential election. In the final three months of the election, fake news that favored either of the two candidates was shared more than 37 million times on Facebook (Allcott and Gentzkow, 2017). Furthermore, most people who are exposed to fake news believe it (Silverman and Singer-Vine, 2016), and a survey by the Pew Research Center shows that "about two-in-three adults say fabricated news stories cause a great deal of confusion about the basic facts of current issues and events" (Barthel et al., 2016). This survey also shows that 23% of surveyed individuals knowingly or unknowingly shared fake news.

Companies like Facebook and Google, which many people depend on as a source of information, have been accused of giving too much prominence to fake news—as a result, these companies are taking steps to prevent fake news sites from collecting advertising revenue (Wingfield et al., 2016). Detecting fake news early can stop a false story before it gains traction on social media, effectively minimizing its influence. An automatic fake news detector could potentially analyze hundreds of thousands of news stories a day and alert readers of stories that are likely to be false or misleading.

2 Proposal

I propose to investigate the use of modern machine learning (ML) and natural language processing (NLP) techniques to automate the process of labeling news stories as "fake" or "real."

2.1 What Is Fake News?

The definition of fake news is subjective (Fake News Challenge, 2017). In order to minimize differences in opinion, it is necessary to choose a narrow definition of fake news. I will adopt the widely used definition discussed in Shu et al. (2017) which defines a fake news story as one that is "intentionally and verifiably false." Fake news must contain information that can be verified as false and must be created with the intent to mislead readers. This definition omits concepts that could be otherwise considered fake news like satire, rumors, conspiracy-theories, and hoaxes.

2.2 Automatic Document Classification for Fake News Detection

The task of automatic document classification is to assign a document to one or more classes or categories. This can be done with a combination of ML to model fake news versus real news and NLP to extract meaningful information from news stories and pre-process text.

Significant research has been conducted on automatic document classification, which will provide me with many benchmarks to explore. Recent efforts in automatic document classification include using binary classifiers like support vector machines (Manevitz and Yousef, 2001) and deep learning techniques (Yang et al., 2016).

2.2.1 Dataset Evaluation and Creation

In order to train an ML model for fake news detection, I will first determine which datasets are best suited for the task. Shu et al. (2017) describe four publicly available datasets which have potential to be used in fake news detection—BuzzFeedNews, LIAR, BS Detector, and CREDBANK.

There is also an ongoing project to develop a reliable dataset for fake news detection on social media called *FakeNewsNet* (Shu et al., 2017). This dataset includes news content and social context features with reliable ground truth fake news labels.

If none of these datasets are well suited to document classification for fake news detection, I will create a new dataset by modifying existing datasets and automatically scraping news stories from the internet. In ML, having a clean and reliable dataset to train models with is of the utmost importance.

2.2.2 Deep Learning

Of particular interest to me is the use of deep learning and neural networks to tackle fake news detection. Advances in computational hardware and data collection have helped neural networks revolutionize tasks like image classification and speech recognition. NLP tasks like part-of-speech tagging and text summarization have also been greatly improved with deep learning, and Yang et al. (2016) have found that a hierarchical attention network performs significantly better than previous methods for automatic document classification. The power and flexibility of deep learning leads me to believe that this is the best approach for fake news detection. However, neural networks are difficult to train and in general require more data than other techniques.

Ruchansky et al. (2017) have applied deep learning to fake news detection with a focus on incorporating the metadata of a news story into the model. However, the datasets they used did not contain traditional news stories—the data was collected from Twitter and Weibo—and they did not experiment with newer neural network architectures such as attention networks. I propose to apply state-of-the-art deep learning techniques in automatic document classification (e.g. hierarchical attention networks) to a clean, reliable fake news dataset.

3 Technical Significance of the Project

In computer science and ML, automatic document classification is a well-studied problem, and fake news detection might seem like a simple application of document classification. However, because of the subjective nature of reporting, it is difficult to find a reliable dataset of fake news stories that can be used for supervised learning. Additionally, news stories are extremely diverse and unstructured—unlike documents like cooking recipes—which makes fake news detection a difficult modeling problem.

In this project, I will determine a reliable dataset for training machine learning models for fake news detection. If no such dataset is publicly available, I plan to create one by modifying existing news datasets and automatically scraping news stories from the web. This dataset will be valuable for researchers in the field who want to experiment with modeling fake news.

The ML and NLP techniques that I plan to experiment with and develop will improve the accuracy and reliability of automatic fake news detection. Ultimately, I will create a useful fake news detector that can assist people in their daily consumption of media.

4 Proposed Timeline

- Fall Quarter 2017
  - Evaluate publicly available fake news datasets—is the dataset sufficiently large and accurate, and does it match my adopted definition of fake news?
  - Implement baseline automatic document classification and fake news detection models like linear classifiers and simple neural networks.
- Winter Quarter 2018
  - Select or create a clean, reliable fake news dataset.
  - Evaluate baseline models against the fake news dataset.
  - Begin implementing more complex neural networks (e.g. LSTMs, attention networks).
- Spring Quarter 2018
  - Experiment with state-of-the-art deep learning techniques.
  - Explore various ML and NLP optimizations (e.g. pre-processing techniques, training algorithms).
  - Build a useable fake news detector system.

5 Conclusion

A reliable automatic fake news detector would be invaluable to organizations struggling with the spread of intentionally misleading and inaccurate news stories. I hope that reliable fake news detection will increase confidence in legitimate reporting and help inform the public.

The state-of-the-art deep learning techniques in automatic document classification I propose to apply to fake news detection will benchmark this modern approach to fake news detection.

References

Hunt Allcott and Matthew Gentzkow. 2017. *Social media and fake news in the 2016 election*. Technical Report. National Bureau of Economic Research.

Michael Barthel, Amy Mitchell, and Jesse Holcomb. 2016. Many Americans believe fake news is sowing confusion. *Pew Research Center* 15 (2016).

Fake News Challenge. 2017. Fake News Challenge Stage 1 (FNC-1): Stance Detection. (2017). `http://www.fakenewschallenge.org`

Larry M Manevitz and Malik Yousef. 2001. One-class SVMs for document classification. *Journal of Machine Learning Research* 2, Dec (2001), 139–154.

Natali Ruchansky, Sungyong Seo, and Yan Liu. 2017. CSI: A Hybrid Deep Model for Fake News. *arXiv preprint arXiv:1703.06959* (2017).

Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake News Detection on Social Media: A Data Mining Perspective. (2017). arXiv:1708.01967

Craig Silverman and Jeremy Singer-Vine. 2016. Most Americans who see fake news believe it, new survey says. *BuzzFeed News (www. buzzfeed. com). URL: https://www.buzzfeed.com/craigsilverman/fake-newssurvey* (2016).

Nick Wingfield, Mike Isaac, and Katie Benner. 2016. Google and Facebook take aim at fake news sites. *The New York Times* 11 (2016).

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alexander J Smola, and Eduard H Hovy. 2016. Hierarchical Attention Networks for Document Classification.. In *HLT-NAACL*. 1480–1489.