

Detecting Fake News from News Articles

Course Project Proposal for CS690N–Advance NLP

Dilip C. Kavarthapu, Manish Motwani, Nishant Yadav

Abstract

Recent studies have shown the decrease of confidence in the mainstream media which indicates that the amount of fake news in such reliable sources is continuously increasing. While much research has been done on detecting fake news from social media posts, little work is done on detecting fake news from the actual news articles. In this project we attempt to automatically detect the fake news articles present on reliable news sources such as `abcnews.com`.

1 Introduction

In news production, dissemination, and consumption, there are ample opportunities to deceive and be deceived. Direct falsifications such as journalistic fraud or social media hoaxes pose obvious predicaments. While fake news may be less malicious, it may still mislead inattentive readers. Taken at face value, fake news can intentionally create a false belief in the readers' minds.

In recent years there has been a trend of decreasing confidence in the mainstream media. According to Gallup polls, only 40% of Americans trust their mass media sources to report the news "fully, accurately and fairly" [1] and a similar survey in the UK has shown that the most-read newspapers were also the least-trusted [2]. One effect of this trend has been to drive news readers to rely more heavily on alternative information sources, including blogs and social media, as a means to escape the perceived bias and unreliability of mainstream news [3]. Ironically, this may leave the readers even more susceptible to incomplete, false, or misleading information [4].

The freedom to write and share posts on social media platforms such as Facebook and Twitter enable users to easily create and share the fake news, making these platforms less reliable news sources than the mainstream media. However, decrease of confidence in mainstream media indicates that the amount of fake news in reliable media sources has increased. Further, it is more challenging to detect such fake news from real news articles as the fabricated information uses the same style and format of journalistic reporting as used for the real news. For example, the fake news headline *Germany Forming EU Super Army Prepar-*

ing For World War 3 With Russia (`pakalertpress.com`) is very similar to the real news headline *World War 3 warning: North Korea COMPLETES nuclear reactor and tests have STARTED* (`www.express.co.uk`). In this project we intend to devise a method that can automatically detect such fake news. We plan to use datasets that comprise of real and fake news extracted from different news sources to train and test our method.

The remaining of this proposal is organized as follows. Section 2 defines the scope of this project, Section 3 defines related work, Section 4 describes the datasets we plan to use in this project, and Section 5 describes the initial set of experiments we plan to conduct.

2 Project Scope

We limit the scope of this project to the problem of detecting fake news solely from the description of news articles crawled from mainstream media sources and not from the social media posts.

3 Related Work

The work relevant to this project can be categorized in terms of the broad area of text classification approaches and their application to solve the problem of fake news detection.

Text Classification Classical classification techniques such as Naive Bayes, SVM, Logistic Regression, etc. [5] use probabilistic models (conditioned on linguistic features of text) to classify the text into desired categories. More recent techniques use deep learning based architectures such as using RNNs and CNNs to encode the features and then use this encoding to perform classification. For example, Kim [6] uses CNNs on sentences with predefined word embeddings for sentiment analysis and question classification.

Fake news detection Traditional fake news identification approaches typically use features such as the linguistic features of the text, source of the text and user responses

on the published text. Ruchansky et al. [7], use scores from the output of the LSTM running over articles and scores from the output of a neural network taking user representations as inputs to make a prediction of the fakeness. Wang [8] compare the performances of various classifiers like SVM, Logistic Regression, Bi-LSTMs, CNNs, etc. to solve this problem. Shu et al. [9] present a comprehensive review of detecting fake news on social media, including fake news characterizations on psychology and social theories, existing algorithms from a data mining perspective, evaluation metrics and representative datasets.

4 Datasets

We plan to use following two publicly available datasets:

- LIAR Dataset [8] : It consists of 12.8K human labeled short sentences from polifact.com, and each statement is evaluated by polifact.com's editor for its truthfulness. Each sentence has one of the 6 fine grained truthfulness labels: *pants-fire*, *false*, *barely-true*, *half-true*, *mostly-true*, and *true*. The dataset also includes meta data such as the speaker, context, affiliation of speaker etc. This dataset is already split into training, validation and test datasets.
- Getting Real About Fake News [10]: The dataset contains headline, text body, and meta-data from 244 websites and represents 12,999 posts in total. The data was pulled using the webhose.io API. Each website was labeled according to the BS Detector [11]. We intend to use this dataset for testing purposes.

5 Proposed Experiments

Wang [8] casts fake news identification as a 6-way classification problem and use Bi-LSTM, and CNN based models to perform this classification. The paper reports that using meta-data together with the short description improves the classifier's accuracy.

In the initial set of experiments, we would implement Bi-LSTM and CNN based models together with pre-trained word2vec [12] word embeddings, train it on LIAR dataset and try to reproduce the results presented in [8]. Our next set of experiments would be to evaluate the trained model on the Kaggle fake news dataset [10] and understand the effect of incorporating additional features of the news on model's performance.

References

[1] R. Riffkin. (2015) Americans' trust in media remains at historical low. [gallup](http://www.gallup.com). [Online]. Available: <http://www.gallup.com/poll/185927/americans-trust-media-remains-historical-low.aspx>

[2] . N. R. Reilly, R. Power, principles and the press. [Online]. Available: <http://www.theopen-road.com/wp-content/uploads/2012/09/Power-principles-and-the-press-Open-Road-and-Populus1.pdf>

[3] Y. Tsftati, "Online news exposure and trust in the mainstream media: Exploring possible associations," *American Behavioral Scientist*, vol. 54, no. 1, pp. 22–42, 2010.

[4] D. Mocanu, L. Rossi, Q. Zhang, M. Karsai, and W. Quattrociocchi, "Collective attention in the age of (mis) information," *Computers in Human Behavior*, vol. 51, pp. 1198–1204, 2015.

[5] R. R. Larson, "Introduction to information retrieval," 2010.

[6] Y. Kim, "Convolutional neural networks for sentence classification," *CoRR*, vol. abs/1408.5882, 2014. [Online]. Available: <http://arxiv.org/abs/1408.5882>

[7] N. Ruchansky, S. Seo, and Y. Liu, "Csi: A hybrid deep model for fake news detection," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, ser. CIKM '17. New York, NY, USA: ACM, 2017, pp. 797–806. [Online]. Available: <http://doi.acm.org/10.1145/3132847.3132877>

[8] W. Y. Wang, "'liar, liar pants on fire': A new benchmark dataset for fake news detection," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, vol. 2, 2017, pp. 422–426.

[9] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *SIGKDD Explor. Newsl.*, vol. 19, no. 1, pp. 22–36, Sep. 2017. [Online]. Available: <http://doi.acm.org/10.1145/3137597.3137600>

[10] M. Risdal. (2017) Getting real about fake news, text & metadata from fake & biased news sources around the web. [Online]. Available: <https://www.kaggle.com/mrisdal/fake-news>

[11] (2017) B s detector. [Online]. Available: <https://github.com/bs-detector/bs-detector>

[12] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.