

Author: Jacob Fuehne

Date: 11/1/2020

Research Paper Summary

Capturing the Style of Fake News, *In the Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)*, 2020

1. Problem definition and the main ideas of the research

Fake news and the issue of credibility in online media has garnered considerable attention lately. Manual fact checking of fact news is laborious and unable to be applied to every post on social media. In this paper, the authors seek to use text analysis to detect documents of low credibility, such as fake news, based on the style that the text is written. To accomplish this task, a corpus of over 103,219 documents from 223 online sources labelled by experts is created and two classifiers are tested on the data.

2. Significance of research study (Importance and Challenges of research problem)

With the viral spread of information on social media, misinformation can be shared rampantly. Manual fact checkers cannot keep up with misinformation, and so algorithms to combat disinformation and tackle the issue of credibility are necessary. An issue in having a classifier for low credibility documents and fake news is that to truly verify a document, a classifier would need a level of text understanding and domain knowledge that might not be sufficient or available. To mitigate this challenge, this paper instead refocuses its efforts on capturing the writing style of low credibility and fake news documents.

3. Main research questions and assumptions

As mentioned, the authors seek to test a methodology for detecting low credibility documents and fake news in online media. Due to the challenges of making a classifier that uses

text understanding, the methods proposed in this paper classify the writing style of the document and compare it to other sources of low credibility and fake news documents. The authors hypothesize that the sources of fake news will typically attempt to attract attention for short term financial or political gains rather than attempt to establish credibility, and as such, they will use informal, sensational, and affective language.

4. Research Methodology

To train the classifiers, the authors create a text corpus of 103,219 documents from 223 sources labelled for credibility based on studies by PolitiFact and Pew Research Center. To gather the corpus, they use a web crawler and convert the HTML pages into plain text using a set of rules. Using this corpus, they train and test Bag of Words and BERT as baseline algorithms and bidirectional LSTM and Stylometric classifier as the style capture models. The baseline algorithms simply perform classifications based on the training data, while the biLSTM and Stylometric classifier classify based on aspects of the style.

5. Experiments

When evaluating these classifiers, three types of tests are performed, document based, topic based, and source based, all with 5 fold cross validation. The document based test includes random documents in the dataset. The topic based test is done with documents that are associated with topics that were identified through LDA topic modelling. And the source based test includes one source with all of its documents for each fold in the 5 fold cross validation. In testing, it was found that style capturing models performed worse than the baseline models in both document based and topic based tests, but performed better in source based tests. This would lead one to conclude that style should not be used to identify credibility in documents. With the better performance of source based testing, a possible explanation is that the models were capturing the

style of particular credible or noncredible sources, rather than the general style of low credible documents.

6. Discussion

6.1 Important aspects

- Tested models in various situations to allow for better discussion/conclusions

Because the authors tested source based cross validation in addition to document and topic based, they were able to draw conclusions about classifiers performance in the discussion about the usefulness of style capturing for detecting fake news.

6.2 Limitations of the paper

- No direction is given for future work

While the paper does a good job of analyzing the results and offering explanations for the shortcomings and successes of the style capturing models, it does not offer any particular suggestions in the conclusion for future work. Instead, it simply encourages fake news detection based on style as a worthwhile direction in research.

6.3 Questions for presenter

- Can you explain what is going on in Figure 2 and why the ranges shown might be so wide ranging?
- Given that there might not be a generalizable fake news style and that true verification would require text understanding, do you think that style capture is a worthwhile direction in research for fake news detection?