

Jeremy Flint

Evaluating Online Articles for Misinformation using Machine Learning and Systematic Data Collection. G69

Engineering Problem

The goal of this project was to add insight to an article's credibility to streamline the process an everyday consumer goes through when deciding how to interpret the information given. This was accomplished by providing the article's sentiment, its source's bias, related articles on the topic, and a coverage report displaying what parts of the political spectrum was reporting the story.

Procedures Used

The bias feature was tested with 5 different urls. The sentiment feature only needed to be functional, the coverage report was tested with 5 test arrays, and the related article iterations were tested against 3 datasets using 3 models of differing strictness on how similar the articles had to be to qualify.

Data

Almost all other features were able to fulfill their criteria, the program was capable of consistently providing the correct source bias, and sentiment. The coverage report feature was also completely accurate. However, the first iteration of the related articles feature only averaged around 52% and was improved to 72% in the second iteration, still falling short of the expected 90%.

Interpretation

The second iteration that removed proper nouns from the entities proved to be more accurate than the first. Its largest improvement was from a 54% accuracy to 88%. However, the second iteration also returned less articles overall. I think that this problem can be mitigated by expanding the size of the article database.

Conclusion

In conclusion, by expanding the database, and making it update real time, this project can help contribute to combating online misinformation.