Final Capstone Report – Satire Detector
Gabriela Tanumihardja
09/20/2020

Satire is a literary style that criticizes errors, vices, and shortcomings of a person, politics, or society. Satire relies on exaggeration, irony, and humour to expose these traits. Although often humorous, laughter is not satire's main goal. By pointing out these shortcomings, satirical writers aim to make their readers think, and possibly seed change in the topic discussed. Some of the world's most read literature are satirical, such George Orwell's *1984*— a cautionary tale about authoritarian government and propaganda. Another well-known satirical novel is Mark Twain's *Adventures of Huckleberry Finn*, which offers a sharp critique of deep-rooted racism in the Southern United States. While it's relatively easy to parse out satire in novels, it is often more difficult to differentiate satire news articles from mainstream journalism. Contemporary satire has been used in mainstream newspapers since the early 1800's, however, most people are currently exposed to news satire through the web. There are a few purely satirical news websites, such as the Onion, the Beaverton, the Babylon Bee, and the Borowitz Report. It is often difficult to identify satirical content from legitimate content and due to satire's controversial nature, this error could cause some problems. A study published by Garrett et al (2019) surveyed over 800 Americans who shared satirical headlines on their social media. It was found that there is a significant divide in the type of articles people believe depending on their political leaning. Republicans tend to believe articles published by the Babylon Bee (a Conservative-leaning satire website), whereas Democrats tend to believe articles published by the Onion (a Liberal-leaning satire website). While it may not be their intention to divide, satirical news headlines could increase the split between people with opposing political views.

In data science, there have been significant developments in Natural Language Processing (NLP). NLP has been commonly used to predict sentiment, analyze trends, and even find online abuse in social media. While models are getting very good at analyzing simpler language structures, I am interested in exploring if machine learning could parse out more complex language structures, such as humour, sarcasms, and ironies. Some extant research has been done in this topic. West and Horvits (2019) created a website, www.unfun.me, that takes real satirical headlines and challenges their visitors to make the headlines 'un-fun' in as few edits as possible. West and Horvits obtained almost 3,000 'un-funned' headlines and utilized machine learning techniques to parse out what makes humour humorous. There also have been studies that focus on detecting satire in the news. Stöckl (2018) utilized linear SVM and logistic regression to identify satire in the news. Stöckl used 60,000 headlines to train his model and was able to obtain remarkable accuracies. For this project, I hope to scrape newer satirical and legitimate news sources to train models that could identify satirical headlines. I would also like to augment the model further, to identify sources of the news headlines. There has been a lot of changes in the world in the past couple of years, sometimes making us feel that we live in a satirical world. It would be very interesting to see if machine learning models still could differentiate the headlines. At the end of this project, I wish to build an API where a custom headline could be entered, and a prediction could be produced.

In starting this project, headlines data were obtained from four different news sources - the Onion, the Beaverton, the Globe and Mail, and the New York Times. Selenium python package was used for browser automation and BeautifulSoup python package was used to scrape necessary data. All of the data obtained were categorical data. Prior to cleaning, over 50,000 headlines were obtained from the four news sources. During cleaning, 9,506 duplicated rows were dropped, as well as 322 rows containing null values. Most of these dropped rows belonged to the New York Times. A target column containing coded information of the news type was added: 0 for legitimate and 1 for satire. For the multilabel classification, a column which codes for the source of the headlines was also added (0 for the Beaverton, 1 for the Globe and Mail, 2 for the Onion, and 3 for the New York Times). An analysis was performed, showing that the two target classes were balanced.

To this dataset, a few preliminary models such as a logistic regression, a decision tree, a KNN, an SVM, a random forest, and an AdaBoost model were fit. Evaluations of the initial logistic regression model reveals that a potential problem with the data. The list of the most predictive words for satirical news include a very biased collection of words such as ‘onion’ and ‘2013’. It also included the word ‘horoscope’, which shows a strong bias for the Onion, as they publish a weekly horoscope article. To alleviate these problems, a resampling of the data based on the article’s published year was called for. An edit to the tokenizer was also performed, allowing it to remove any numbers, and the words ‘onion’ and ‘horoscope’. Using this tokenizer, two wordclouds were created, one for each type of article. It was very interesting to see that while there are many common words between the word clouds, there are some distinct patterns as well. Satirical headlines used more words such as ‘man’ and ‘area’, whilst legitimate headlines report more on the coronavirus pandemic.

Using this more balanced dataset, logistic regression, SVM, Bayes, AdaBoost, XGBoost, and random forest models were fit. Validation accuracy scores ranging from 63% (XGBoost) to 79% (logistic regression) were obtained. The most predictive words coming out of the newer logistic regression model appeared to be less biased. In line with the wordclouds produced earlier, legitimate news appeared to cover more pandemic news, whereas satirical news’ most predictive words include ‘man’, ‘area’, and ‘local’. This is somewhat expected, as satirical headlines often begin with ‘Area Man’ and ‘Local Man’. While I was satisfied with the interpretability and simplicity of the logistic regression, I wanted to try a deep learning for this problem. A transfer learning method using BERT, Google’s state-of-the-art NLP model, was performed (Devlin et al., 2019). Google’s (2019) sample code was altered to fit this specific classification problem. From this model, a very high test-accuracy of 90% was achieved. A sample of misclassified headlines were extracted. In analyzing these headlines, I found that it would’ve been difficult for humans to differentiate these headlines, as a lot of them were very contextual. It would be very interesting to conduct an experiment comparing human's ability to distinguish the satirical headlines and obtain a baseline rate.

In predicting the source of each headline, the data were further resampled, ensuring a balance across the news sources. Once again, Google's (2019) example code was modified to fit the current multi-label classification problem. For this four-label classification task, a test-accuracy of 80% was obtained. From the produced confusion matrix, it could be seen that the model had more problems in identifying articles coming from the Beaverton. A lot of these articles were misclassified as coming from the Onion. This could be explained by the fact that both of them are North America's largest satirical news sources and they tend to focus on similar topics.

From here, a simple app was created using StreamLIt where a novel headline could be entered. If deployed, this simple predictor could be useful as people could check the legitimacy of a news article before believing it. I would also like to add a functionality to this app where it could suggest a similar article of the opposite type. I would also like to add another functionality which shows the headline's most important features. This would be very interesting to see, as it could potentially show determine the key ingredients of satire. I also hope to apply some of the things and ideas I have obtained from these experimentations to another project where the political lean of an article could be detected from its headline.

References

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). Financial Sentiment Analysis for Predicting Direction of Stocks using Bidirectional Encoder Representations from Transformers (BERT) and Deep Learning Models. *ArXiv*. doi:10.17758/uruae8.ul12191013

Garrett, R, & Poulsen, S. Too many people think satirical news is real. Retrieved September 19, 2020, from https://theconversation.com/too-many-people-think-satirical-news-is-real-121666

Google (2019). Movie Reviews with bert-for-tf2.ipynb. Retrieved September 20, 2020, from https://colab.research.google.com/github/bentoml/gallery/blob/master/tensorflow/bert/bert_movie_reviews.ipynb

Stöckl, A. (2018). Detecting Satire in the News with Machine Learning. Retrieved September 20, 2020, from https://www.researchgate.net/publication/328018762_Detecting_Satire_in_the_News_with_Machine_Learning

West, R., & Horvitz, E. (2019). Reverse-Engineering Satire, or "Paper on Computational Humor Accepted despite Making Serious Advances". *Proceedings of the AAAI Conference on Artificial Intelligence, 33*, 7265-7272. doi:10.1609/aaai.v33i01.33017265