SENTIMENT ANALYSIS OF NEWS ARTICLES USING BIDIRECTIONAL RECURRENT NEURAL NETWORKS

ALEXANDER TAN KA JIN

A project report submitted in partial fulfilment of the
requirements for the award of the degree of
Your Degree

Faculty Of Computing
Universiti Teknologi Malaysia

21/12/2024

Introduction

The subject of Natural Language Processing is the application of machines and computation within the domain of language. Part of it is Sentiment Analysis, which analyses neutrality, opinion strength and positivity/negativity from text and language use. Within recent times, the domain has utilized the tools of Deep Learning and Artificial Intelligence in order to analyse the complicated patterns of semantics and syntax, moving away from older machine learning methods and rules-based lexicon methods. The use of models like LSTM and GRU allow better performances in the study of language use in everyday life such as within social media sites like Twitter (now X).(Wankhade et al., 2022) Language can be used objectively or subjectively, either a person is stating facts about a subject or giving their emotional opinions or narratives around the subject. One domain that is often not explored within sentiment analysis is the domain of long-form text like news articles.

Politics is a domain that tends to be filled with various opinions within a given subject. Whether a given policy, law, political figure or institution is good or bad is the main form of discussion within politics. Despite their intention of bringing objective news, journalism and news articles are written by people and institutions with certain political opinions and ideologies and may create narratives that reflect their political biases through the use of language. (Baly et al., 2020) These opinions and ideas are often embedded in otherwise objective journalism and Whether intentional or not, this use of language can push narratives that sways people into certain political beliefs and camps and is inconducive to a proper functioning democratic society. Compared to machine learning models, the use of deep learning models allows machines to extract certain features on their own and allows them to potentially detect more complex patterns of sentiment.

This thesis aims to perform sentiment analysis on news of various forms of bias in order to uncover the relationship between sentiment and propaganda. It uses a deep learning approach in order to analyse this relationship and how subjective viewpoints are pushed through objective language.

Problem Background

Recent studies of Sentiment Analysis (see Literature Review) have increased in their application of Deep Learning models such as LSTM and Transformer Models. This is a trend that is unique to contemporary times where the internet is adopted by billions of people and millions of data are made readily available, which is perfect for the use of large data deep learning models. Often the subject and dataset of analysis is short-form media like Twitter (now X) posts and consumer opinions, less work is done on long form media such as news articles.

The main issue of sentiment analysis in news is the tendency for journalists to phrase words objectively rather than utilizing subjective language. (Alonso et al., 2021) This provide a unique challenge for sentiment analysis as sentiment on news articles is expressed more subtly compared to messages on social media. (Balahur et al., 2013) Lexicon-based techniques rely weighing the importance of certain words and analysing the sentiment of each word which can fail to pick up the finer details of propaganda in news. (Da San Martino et al., 2019) identified 18 methods of language use in propaganda within news articles which are often overlooked by classification models and techniques, the same article also notes that previous studies on propaganda in news tended to label sources as biases and not news.

Due to the recent success in identifying patterns within text using deep learning, deep learning models such as Long Short Term Memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT) provide promising avenues for sentiment analysis, especially on text that contains more subtle forms of sentiment. The consideration of these insights from previous studies provide an avenue for investigating the correlation between sentiment and political propaganda and therefore political biases within news outlets and articles.

Problem Statement

This thesis aims to perform an analysis of the relationship between sentiment within text and use of language within news using deep learning models. The research aims to address the lack of sentiment analysis on long-form text data that is published within various media outlets rather than focusing on short-form text data within social media. In particular it aims to uncover and analyse in detail how subjective opinions are embedded within objective text. It analyses the efficacy of sentiment analysis with regards to identifying the degree of opinionated news. The analysis will also determine the relationship between media outlets and biased and opinionated journalism. With the use of Deep Learning technology, this research aims to use models such as LSTMs to better analyse textual data. This research aims to help citizens understand the use of language better within media and helps understanding how journalism can manipulate language to advocate or disagree with certain figures and policies.

Research Objectives

1. Perform an exploratory sentiment analysis on large news datasets to identify patterns and trends in the language use of journalism.
2. Construct BiLSTM and BiGRU models to uncover complex language patterns and potentially reveal sentiment hidden under objective language.
3. Conduct comprehensive evaluation on the model to ensure it's performance.
4. Deploy the model onto a large dataset and visualize article sentiment by each media outlet.

Research Questions

1. What is the efficacy of utilizing Deep Learning models on long-form text data compared to Machine Learning/Lexicon based sentiment analysis?

2. To what degree does sentiment analysis properly account for the analysis of propaganda within news outlets and articles?
3. Is there a relationship between new media outlet popularity, bias and degree of opinionated language?

Significance of the Research

The research can help people identify the biases of new sources and how sentiment and opinions are used as forms of propaganda. Which can help people critique media institutions and ensure that people can differentiate opinion and facts within a long article. Additionally, the research contributes to the use of deep learning models within sentiment analysis of long-form textual data and sentiment analysis of objective and neutral language.

Study Scope

The study uses large sets of data from recent time and only considers the use of models, BiLSTM and BiGRU. The research will conduct preliminary analysis of the data, extensive data cleaning and preprocessing before it is used to train our choice of model. The model will then be meticulously evaluated for it's robustness and performance before being used to predict emotion sentiment of articles. The scope of the analysis of articles will cover the purported media bias of it's original institution and it's popularity as a source of news along with the sentimental score of the article. All of it will be presented in the study as a visualization using graphs. Finally, the project will examine and address it's limitations for future work and enhancements in both the use of data and the construction of the model.

References

Alonso, M. A., Vilares, D., Gómez-Rodríguez, C., & Vilares, J. (2021). Sentiment analysis for fake news detection. *Electronics*, *10*(11), 1348.

Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Van Der Goot, E., Halkia, M., ... Belyaeva, J. (2013). Sentiment analysis in the news. *arXiv preprint arXiv:1309.6202*.

Baly, R., Martino, G. D. S., Glass, J., & Nakov, P. (2020). We can detect your bias: Predicting the political ideology of news articles. *arXiv preprint arXiv:2010.05338*.

Da San Martino, G., Seunghak, Y., Barrón-Cedeno, A., Petrov, R., Nakov, P., et al. (2019). Fine-grained analysis of propaganda in news article. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (emnlp-ijcnlp)* (pp. 5636–5646).

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*(7), 5731–5780.