Advanced NLP - Final Project Report

Bias Detection in News

Team-10: Coruscant

SAHIL BHATT (2018111002) SRINATH NAIR (2018111001)

Introduction

The goal of this task is to detect bias in news articles. It is a binary classification problem that outputs whether or not the content in the news is biased. The biased or hyper-partisan reporting of news is done in a way that strongly favours one position (mostly political) and would be in fierce disagreement with the opponents. Hyperpartisan news reporting often involves either stretching the truth or breaking it with fake news and are often spread quickly due to its highly sensational content. The task is to detect this hyperpartisan language in news articles.

Related Work

Team Bertha-von-Suttner in the SemEval 2019 task 4 Hyperpartisan News Detection task used sentence representations from averaged word embeddings generated from the pre-trained ELMo model with Convolutional Neural Networks and Batch Normalization for predicting hyperpartisan news. The final predictions were generated from the averaged predictions of an ensemble of models. With this architecture, their system ranked first place, based on accuracy, the official scoring metric.

Team Vernon-Fenwick from the Samsung R&D was behind the team ranked first place by mere 0.16 points in the shared task SemEval2019 task 4. The team used a concatenated set of Handcrafted features (HF) and Semantic features to obtain the article representations. They tried a combination of Doc2vec + HF, Glove + HF and Universal Sentence Encoder (USE) + HF and obtained the best accuracy of 82.01 using the USE + HF combination.

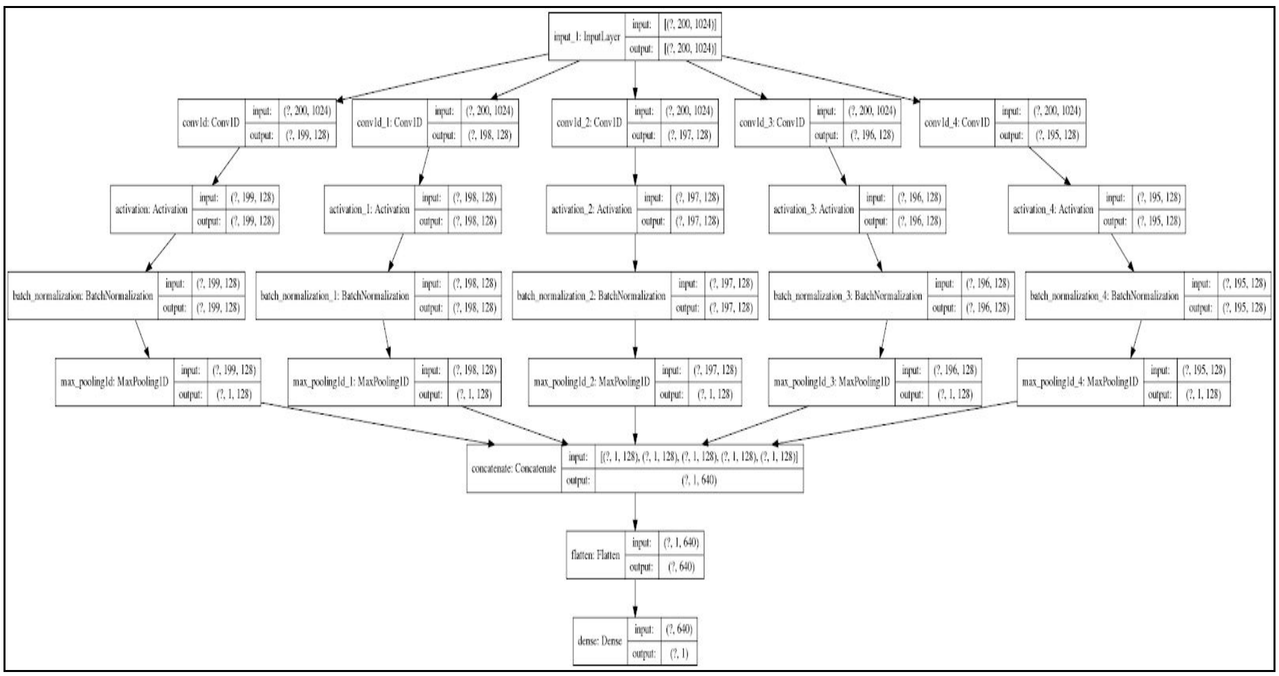
Martin Potthast et al. (2017) used a corpus of 1627 fact checked articles containing both hyperpartisan news from the left-wing and the right-wing, and mainstream publishers. The work tried to find the similarities in writing style of the left and right wing publishers. The study revealed that the style of writing of left-wing and right-wing news have a lot more in common than any of the two have with the mainstream. Furthermore, they showed that hyperpartisan news can be distinguished well by its style from the mainstream (F1 = 0.78).

Model Architecture

Our baseline paper, whose performance we tried to replicate, used a convolutional neural network (CNN) to train the classifier. The inputs to the CNN were ELMo embeddings generated using AllenNLP. This baseline model used 5 convolutional layers followed by a ReLU activation function. This was followed by batch normalization and max-pooling of the output. The outputs were then combined to form an input to a fully connected layer to get a single output. A sigmoid function is then employed for the binary classification task.

We were able to replicate the performance of the baseline by using the same architecture as theirs although we used the Simple ELMo package to generate the embeddings since ELMo was found to be deprecated in AllenNLP.

We then proceeded to generate BERT embeddings for the by-article corpus which was fed into the CNN to get improved results over the baseline.



Preprocessing Techniques

The dataset is a crowdsourced set of 1273 articles annotated manually by 3 annotators and contains political news. Out of these 1273 articles, 645 were used as the training dataset. The test dataset was made private by the organisers of the shared task. So we decided to split the training dataset of 645 articles available publicly into 80-20 train data and test data.
Our input file was in XML format. The preprocessing steps involved parsing this file, filtering the text using spaCy, replacing XML tags, and then adding article id, bias, title and articles line-by-line to the output file.
A sample article in the XML file would look as follows:

<**article id**="0182515" **published-at**="2007-01-22" **title**="They're crumbling">
<**p**>What a pleasant surprise to see Jacques Leslie, a journalist and real expert on dams, with a long <**a href**="http://www.nytimes.com/2007/01/22/opinion/22leslie.2.html

?ex=1327122000&en=42caf99f05e4cba8&ei=5090&partner=rssuserland&emc=rss" **type**="external">op-ed<**/a**> on the hallowed pages of the New York Times. Leslie, author of <**a href**="" **type**="internal">Deep Water: The Epic Struggle Over Dams, Displaced People and the Environment<**/a**>, highlights the threat posed by poorly maintained and increasingly failing dams around the country:<**/p**>
<**p**>Unlike, say, waterways and sanitation plants, a majority of dams - 56 percent of those.. <**/p**>
<**p**>Kinda makes you want to find out what is upstream.<**/p**> <**/article**>

The input file was stored in a TSV format, which was then used to generate embeddings using ELMo and S-BERT.

ELMo

ELMo is a deep contextualized word representation that models (1) the complex characteristics of word use and (2) how these uses vary across linguistic contexts. These word vectors are learned functions of the internal states of a deep bidirectional language model (biLM), which is pre-trained on a large text corpus. They can be easily added to existing models and significantly improve the state of the art across a broad range of challenging NLP problems, including question answering, textual entailment, and sentiment analysis.

BERT

The BERT model was proposed in BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding by Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. It’s a bidirectional transformer pre-trained using a combination of masked language modelling objective and next sentence prediction on a large corpus comprising the Toronto Book Corpus and Wikipedia. The pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as question answering and language inference, without substantial task-specific architecture modifications.

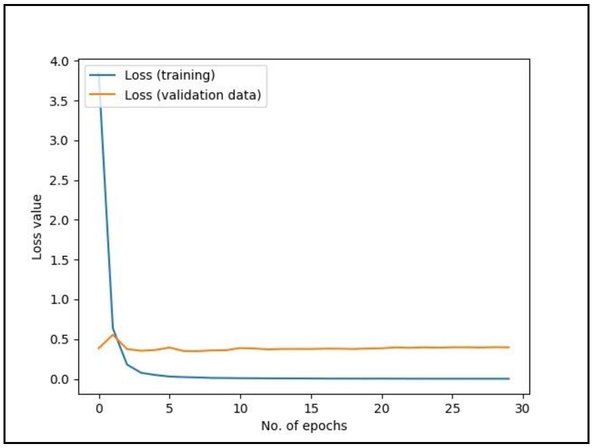
S-BERT

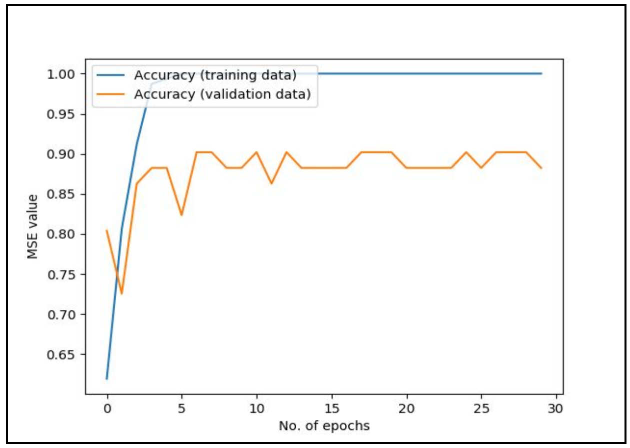
S-BERT or Sentence BERT is a modification of the pre-trained BERT network, which uses Siamese and triplet network structures to give sentence-level embeddings that can be compared using cosine similarity.

Results and Observations

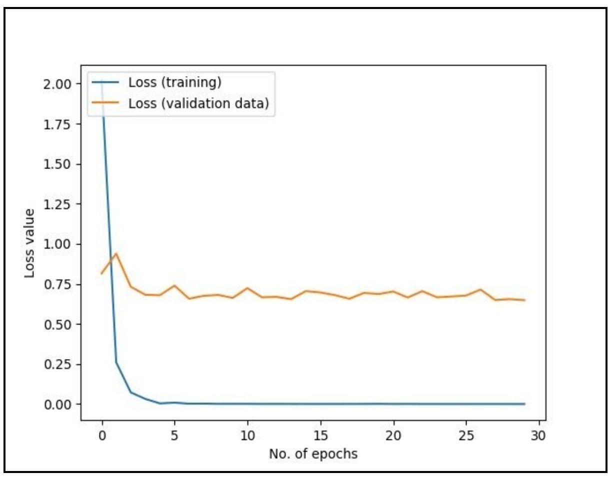
We trained our models on both the ELMo and S-BERT embeddings and obtained the following training and validation loss and accuracy plots.

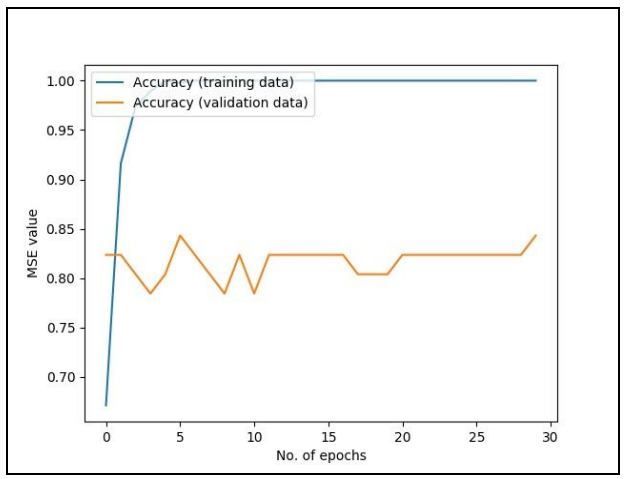
ELMo





S-BERT





On the 129 articles used as our testing set, we observe an F1-score of 0.752 when we used our CNN model trained with ELMo embeddings. With S-BERT embeddings, we observed a significant improvement, as it yielded an F1-score of 0.822.

We also devised a new metric to compare our models by measuring average confidence scores of classification and misclassification.

These average confidence scores were computed by taking the mean of all sigmoid outputs for our predictions - both for correct and incorrect classifications.

For ELMo, the average confidence in the correct classification was 0.845, while for misclassification it was 0.741.

For S-BERT, the average confidence in the correct classification was 0.892, while for misclassification it was 0.749.

MODEL	***F1-score***	***Correct classification confidence***	***Incorrect classification confidence***
ELMo	0.752	0.845	0.741
S-BERT	**0.822**	0.892	0.749

Qualitative Analysis

Our training dataset had a mix of biased and unbiased articles. We also observe (by glancing over the dataset), that a majority of the articles in consideration pertain to the 2016 US elections.

Examples of articles (shortened) that were labelled as hyper-partisan are as follows:

1. *Donald Trump ran on many braggadocios and largely unrealistic campaign promises…Trump Just Woke Up & Viciously Attacked Puerto Ricans On Twitter Like A Cruel Old Man.*
2. *Politicians are trying to deceive you. Modern (American) politics is essentially a game in which politicians say and do whatever they possibly can to persuade people that they are an ideal leader deserving of a vote. …*

On the other hand, some of the non-hyper partisan articles (shortened) were as follows:

1. *Hillary Clinton Campaigns In Des Moines As Early Voting Begins In Iowa. The final full day of campaigning is over, and on Tuesday, Americans will determine which candidate will be the next president of the United States. …*
2. *Brazil's new Migration Law relaxes criminal penalties for migrants. ...*

We recognize bias if a statement reflects prejudice, partiality or preference for or against a person, object or idea.
The language in such articles is generally extreme and appeals more to the emotions rather than logic. A limited or one-sided view of the topic is presented to the reader. Since our dataset mostly contains articles taken from various websites, one might even look for other clues such as if the website is trying to 'sell something', promote a certain idea or stands to gain from publishing such articles.

As a sort of sanity check for the ELMo and S-BERT embeddings we generated, we have used t-SNE and PCA to reduce dimensionality and visualize the embeddings on a small subset of the dataset using an interactive map using Plotly (link to interactive map: ELMo , S-BERT). We observe the similarity in the articles close to each other. Darker shades represent shorter sentences and lighter ones represent longer ones.

Some details about our dataset are given below:

Avg number of words : 527 in SBert-Elmo mismatch (27 articles)
Avg number of sentences : 28 in SBert-Elmo mismatch (27 articles)
Avg number of words in articles in the whole dataset : 587
Avg number of sentences in articles in the whole dataset : 32
Length of longest article(s) : 257 sentences
Length of shortest article(s) : 1 sentence
(**Note**: *SBert-Elmo mismatch means the number of articles where Elmo predicted true but Sbert predicted false, and vice-versa*)

We have also plotted graphs trying to correlate the classification and misclassification probabilities versus the length of the article, but did not observe any meaningful relation between them. The article length did not seem to have any effect on how it was eventually classified by both the ELMo and SBERT models.

Improvements and Future Work

The CNN that is being used to train the classifier can be replaced with some other Neural Network. We would recommend starting off by trying out RNNs which are used often in NLP tasks.

We would also suggest training with a larger dataset. As far as a generalization of our model goes, we have to keep in mind that the dataset used for training our model was handpicked to be political in nature. The larger dataset that could be used can have more generic articles to help generalize better to all articles that could be input to the model.

Furthermore, this task can be redone as a multiclass classification problem where the model can be trained to distinguish between left leaning articles, right leaning articles and unbiased articles.

Bibliography

Deep contextualized word representations (Matthew et al.): https://arxiv.org/pdf/1802.05365.pdf

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (Jacob et al.): https://arxiv.org/pdf/1810.04805.pdf

SemEval-2019 Task 4: Hyperpartisan News Detection: https://www.aclweb.org/anthology/S19-2145.pdf

Team Bertha von Suttner at SemEval-2019 Task 4: Hyperpartisan News Detection using ELMo Sentence Representation Convolutional Network (Jian et al.): https://www.aclweb.org/anthology/S19-2146/

Vernon-fenwick at SemEval-2019 Task 4: Hyperpartisan News Detection using Lexical and Semantic Features (Srivastava et al.): https://www.aclweb.org/anthology/S19-2189.pdf

The Sally Smedley Hyperpartisan News Detector at SemEval-2019 Task 4: Learning Classifiers with Feature Combinations and Ensembling (Hanawa et al.): https://www.aclweb.org/anthology/S19-2185.pdf

A Stylometric Inquiry into Hyperpartisan and Fake News (Potthast et al.): https://arxiv.org/abs/1702.05638

Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks: https://www.aclweb.org/anthology/D19-1410.pdf