Aalto University
School of Science

Measuring Political Bias in British Media:

Using Recurrent Neural Networks for Long Form Textual Analysis

Rory How

May 29, 2020

1

Introduction

1 Word of warning

- This presentation is *slightly* heavy
- This is about as light as I can make while preserving what I *actually* did
- I promise I'll try and keep it as light and as speedy as possible
- There's some formulas but I promise you don't need to know what they do, only that they exist

1 Introduction

We are all consuming media on a daily basis. This is almost entirely unavoidable.
We, the consumers, have a certain level of trust that the media supplied to us is truthful, reliable, and valid.

1 The 2016 UK Referendum of EU Membership

- On 23rd June 2016, The British public voted to leave the EU, with 51.89% voting to leave, and 48.11% voting to remain.
- Much of the the British media gave explicit endorsements to the pro-leave or pro-remain campaigns, respectively. [5]

Is there a way that we can use these technologies to predict political bias in the British print media?

1 Research Questions

- Are we able to find a way in which to determine political bias in the traditional British media?
- Are we able to find a way in which to predict a political bias in supposedly unbiased outlets, such as the BBC?
- Which machine learning models produce the highest amount of accuracy and shortest training time in which to make effective predictions of the political biases of news articles?

2

Background

2 The British Public Believe in Media Bias

In 2019, 2040 British Adults were asked the following question:

> *"Some people talk about 'left', 'right' and 'centre' to describe parties and politicians. With this in mind, where would you place each of the following?" [3]*

The results show that much of the British public has quite a strong unity in terms of describing the partisan nature of the British traditional print media.

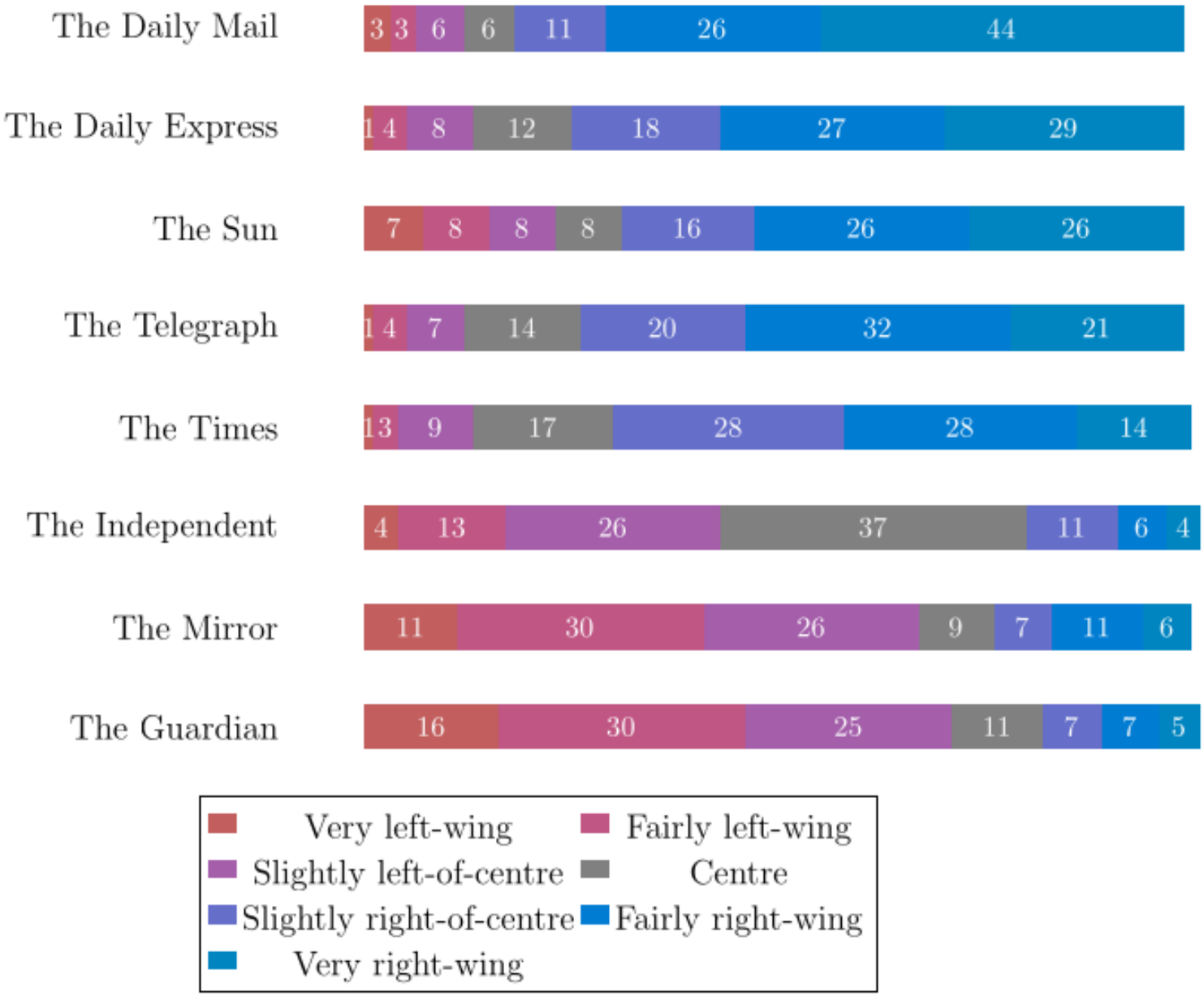

The Daily Mail
3 3 6 6 11 26 44
The Daily Express
1 4 8 12 18 27 29
The Sun
7 8 8 8 16 26 26
The Telegraph
1 4 7 14 20 32 21
The Times
1 3 9 17 28 28 14
The Independent
4 13 26 37 11 6 4
The Mirror
11 30 26 9 7 11 6
The Guardian
16 30 25 11 7 7 5
Very left-wing
Fairly left-wing
Slightly left-of-centre
Centre
Slightly right-of-centre
Fairly right-wing
Very right-wing

Figure: A side-by-side contrast of headlines published by a pro-leave and pro-remain outlet relating to the referendum vote, respectively.

2 The British Media are Consistent in Endorsements

- Traditionally, the British newspapers will explicitly give endorsements to certain campaigns surrounding a referendum or general election
- Many of the newspapers give endorsements that align with certain political philosophies, and they tend to stay true to these political philosophies over the course of many years.
- Typically, the Conservative Party gain the most support from the British press surrounding these crucial votes.

Newspaper	2010 GE	2015 GE	2017 GE	2019 Circulation
The Sun	Conservative	Conservative	Conservative	1 371.19
The Daily Mail	Conservative	Conservative	Conservative	1 199.76
The Mirror	Labour	Labour	Labour	499.82
The Times	Conservative	Conservative	Conservative	406.28
The Telegraph	Conservative	Conservative	Conservative	335.74
The Express	Conservative	Conservative	UKIP	312.77
The Financial Times	Conservative	Conservative	Conservative	168.55
The Guardian	Lib Dem	Labour	Labour	134.57
The Independent	Lib Dem	Lib Dem	None	N/A

Figure: Newspaper endorsements given for general elections (denoted in table as GE) in 2010 [7], 2015 [8] and 2017 [9]. 'None' denotes that the paper made no endorsement for that election. The final column denotes the 2019 circulation for that paper, in thousands [2]. Here we use the statistics for the weekday edition of the newspaper: For example, We use *The Sun*'s daily readership figures as opposed to *The Sun on Sunday*'s.

2 The BBC Impartiality

The state-owned BBC contains in its editorial guidelines [1],
a statement declaring its complete impartiality in all content created:

> *"The BBC is committed to achieving due impartiality in all its output. This commitment is fundamental to our reputation, our values and the trust of audiences."*

2 Types of Bias

- **Selection bias** is when a story is not covered by a certain outlet at all i.e. certain stories are selected to be covered, while others are omitted.
- **Coverage bias** is when a story can occupy more physical space on a newspaper, or occupy a large / higher spot on a website.
- **Framing bias** is when facts are conveyed in such a way to steer the audiences opinion in a certain direction.
- **Statement bias** is the perspective of the individual who is writing the article. These are the opinionated comments that reflect the authors beliefs.

2 Naive Bayes

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{1}$$

Intuitively, the Bayes Theorem states that we can find the probability of A happening given the occurrence of B. To put this into the perspective of a text classification problem, we can find the probability of a sentence $X = (x_0, x_1, \ldots, x_n)$, containing individual words x_i, having a pro-remain bias (i.e a label y),given a similar article B which is known to have a similar bias.

$$P(y|X) = \frac{P(y|X)P(X)}{P(y)}. \tag{2}$$

2 The Multi-layer Perceptron

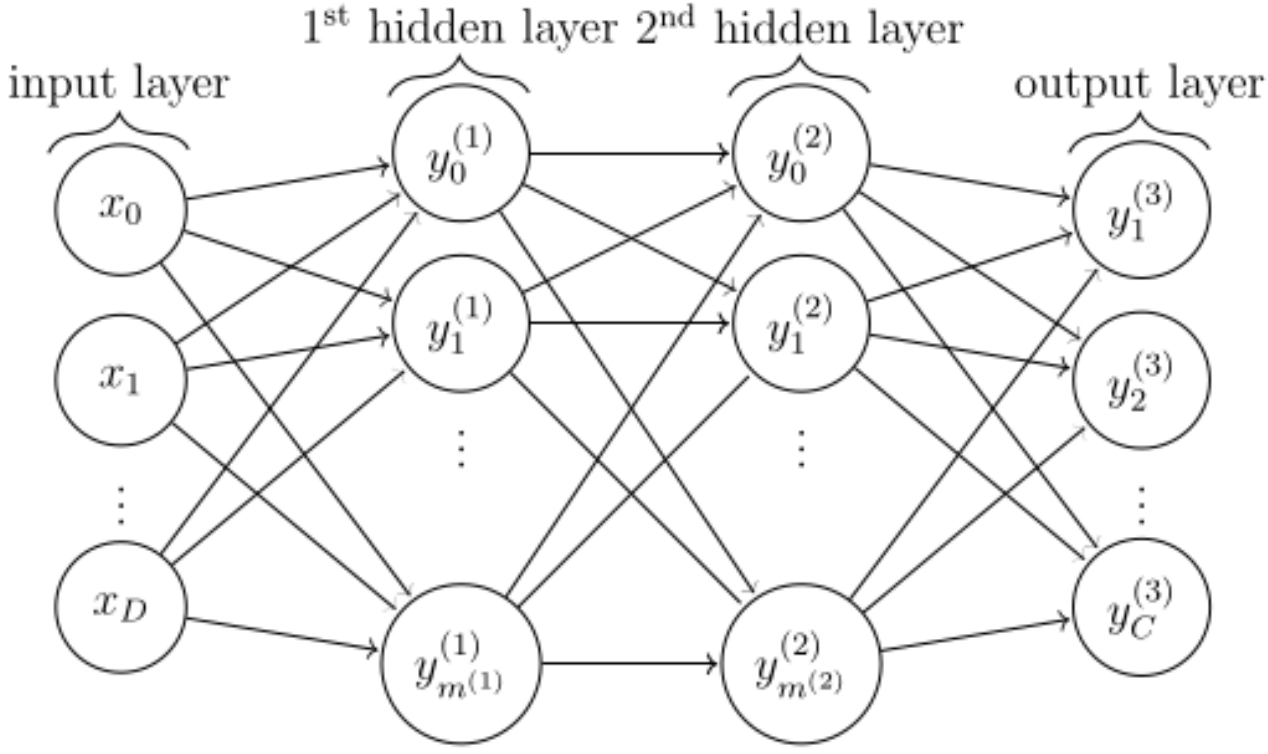


Figure: Network graph of a 3-layer perceptron with D input units and C output units. Each hidden layer l contains $m^{(l)}$ hidden units.

2 The Multi-layer Perceptron

The goal of the the multi-layer perceptron is to estimate some function, f^*. If we are to have a classifier $y = f^*(\boldsymbol{x})$, we are mapping our input, $\boldsymbol{x}$ to our label, $\boldsymbol{y}$. In a typical feed forward network, each node in the perceptron that has a classification rule as follows:

$$f(\boldsymbol{x}) = \begin{cases} 1, & \text{if } \boldsymbol{w} \cdot \boldsymbol{x} + b > 0 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

If we stack linear classifiers on top of each other, with activation functions, we are able to capture nonlinear relationships in data.

2 Recurrent Neural Network Basics

- **Recurrent neural networks (RNNs)** are a type of neural network that can be used to perform classification tasks over sequences of undefined lengths.
- The **Gated Recurrent Unit (GRU)** [4] developed in 2014 utilises a specific architecture to be able to process long sequences has proven popular in more recent times.
- The **Long Short-Term Memory cell** [6] is another approach commonly used for learning behaviours in long sequences of data.

2 Recurrent Neural Networks

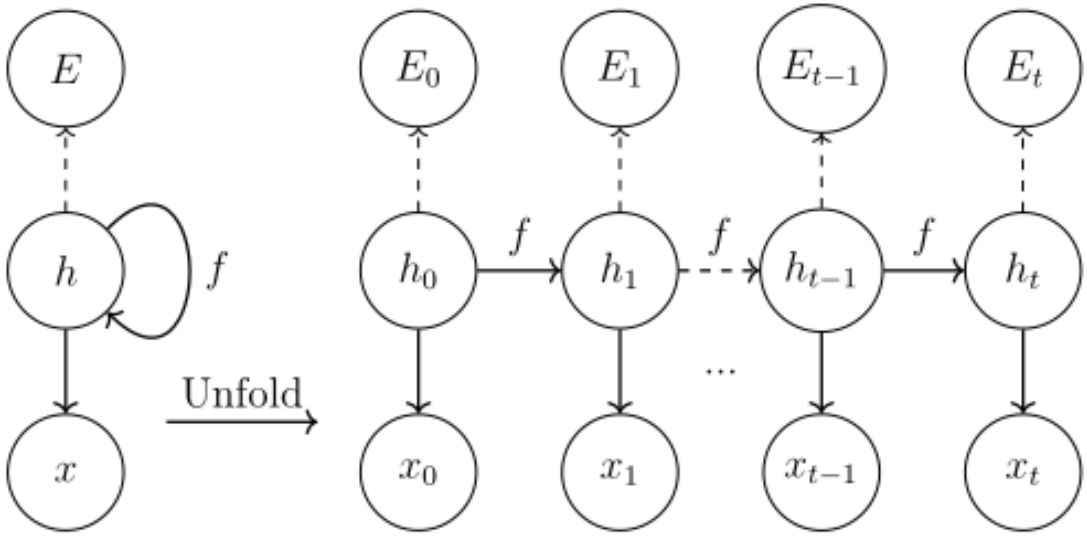


Figure: A basic recurrent network architecture, without any activation layers or outputs. Hidden states are calculated from left to right using the same objective function, taking in new parameters *x* at each time step *t*

2 GRU Cell

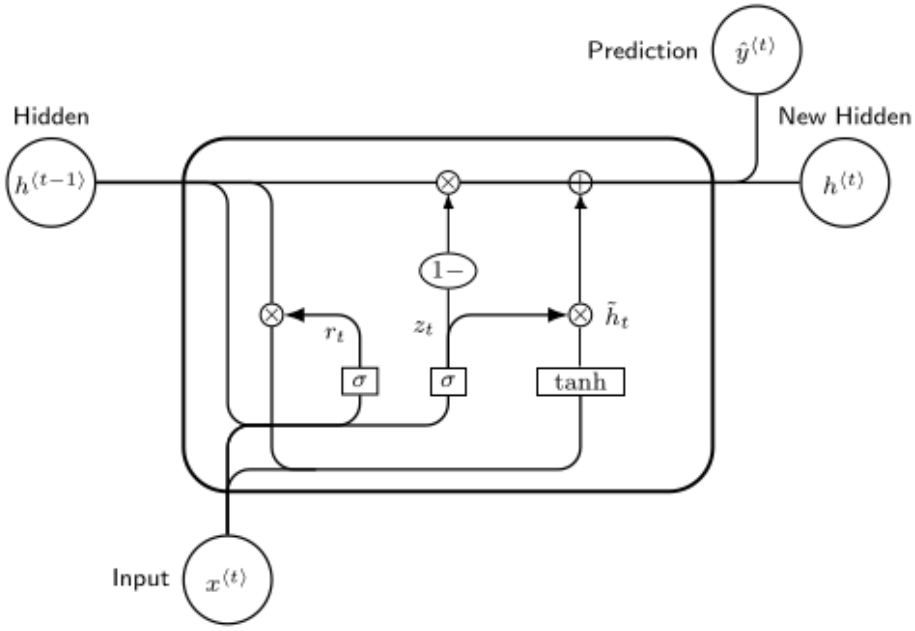


Figure: The GRU (and LSTM) cells utilise “gates” to regulate the flow of data throughout the network. Here we use two sigmoid functions and a tanh activation function to calculate the hidden state h^t

2 Word Embeddings

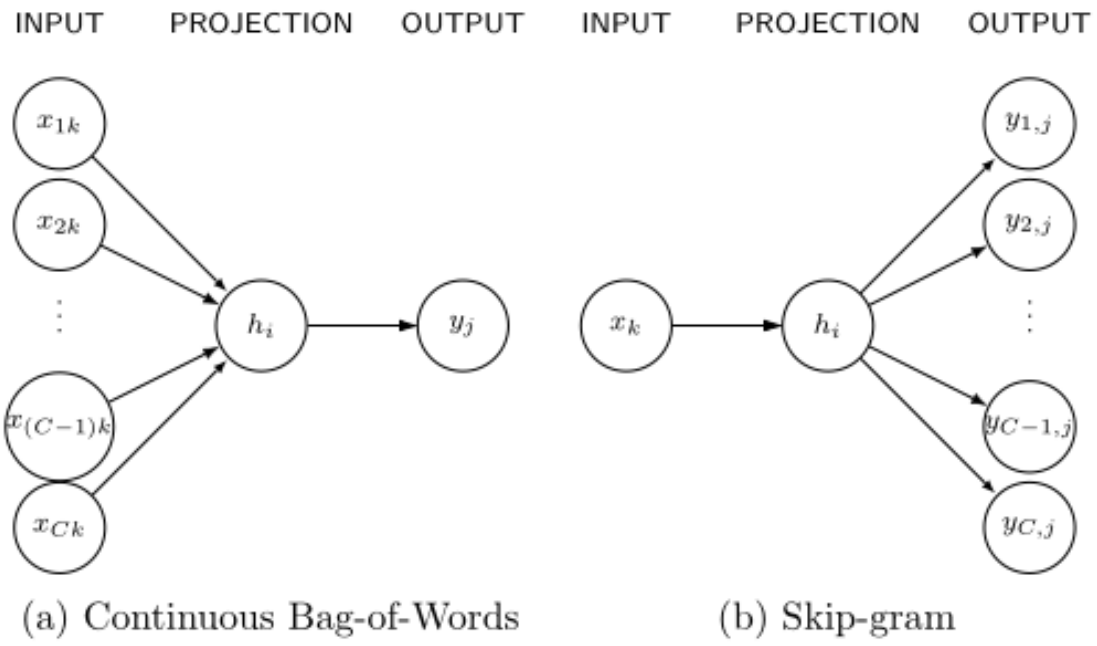


Figure: Model architectures showcasing the intuition between Skip-gram and CBOW methods of Word2Vec.

3

Methods

3 Word Embeddings

- We are able to tune a variety of parameters when training our **Word2Vec** models over our input data.
- These parameters are **window size**, **minimum quantity of word**, **quantity of negative samples used**, **embedding vector size**.
- We can then use common ways of determining the quality of the created embeddings.
- These include the Loss value from training, and **Pearson** / **Spearman** correlation coefficient values

3 Recurrent Neural Networks

- Similarly, we can tune parameters when training our recurrent neural network in order to achieve the best performing model
- These are **batch size**, **hidden layer size**, **dropout percentage**, and **number of RNN layers**.
- We can then use the **F1 Score**, which is a kind of average using **Precision** and **Recall** calculations to determine the quality of our network.

$$\texttt{F1 Score} = 2 \cdot \frac{\texttt{Recall} \cdot \texttt{Precision}}{\texttt{Recall} + \texttt{Precision}}. \tag{4}$$

We also compared the best performing model to a **naive bayes** as a baseline of how the model compares to a more simple approach.

4

Implementation

4 Data Collection

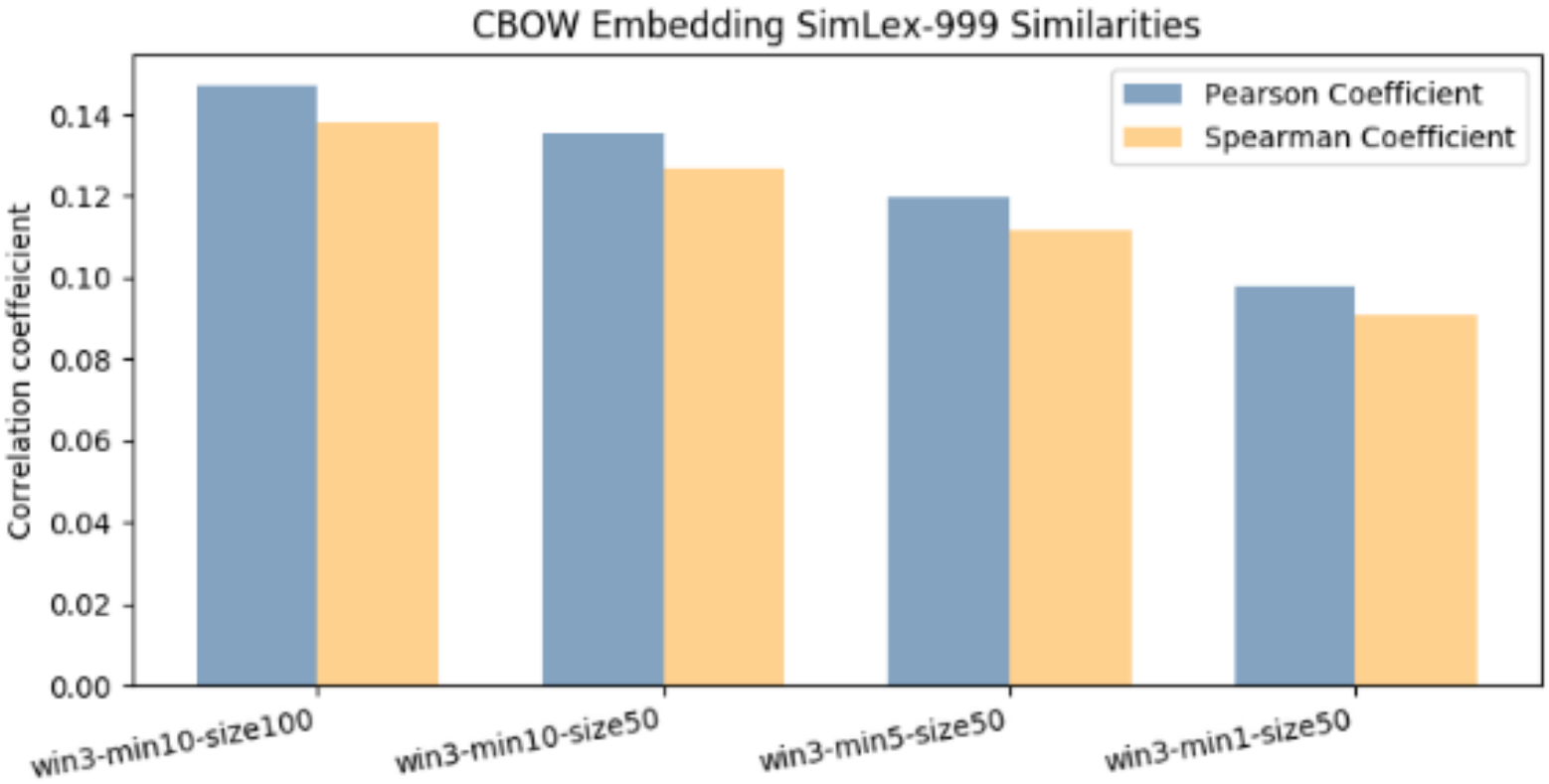
- EU referendum relating articles were retrieved from **Event Registry**, an online service that collects articles from many different sources.
- The articles were published between the dates 2015-06-23 to 2016-06-23
- In total, 5232 articles were retrieved.
- This resulted in a total of 101415 sentences, each of which would be assigned a political leaning based on the explicit endorsements given by the newspaper.

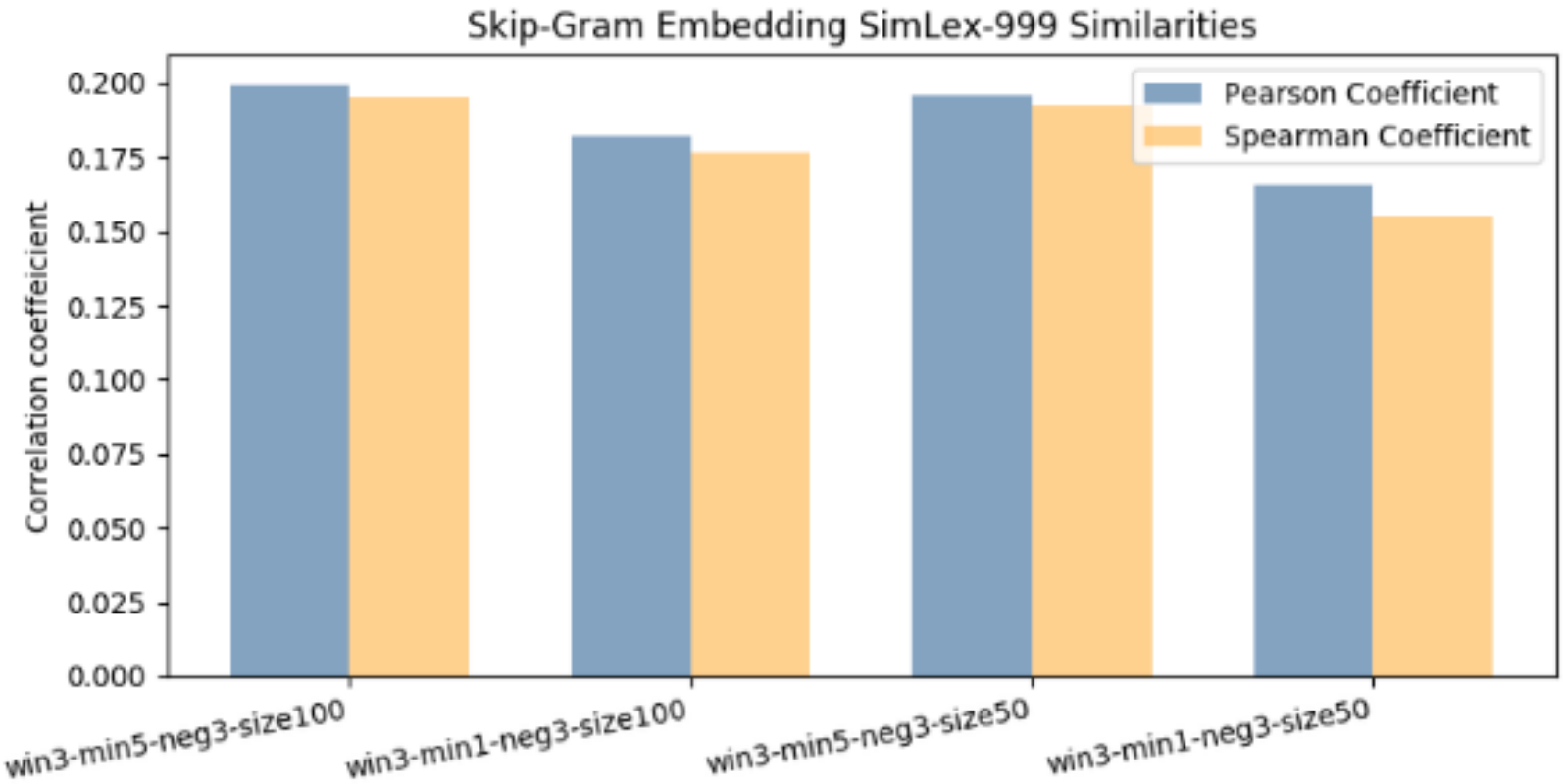
4 Tools and Frameworks

- The word embeddings were created using a text-processing framework called **Gensim**.
- **Pytorch**, a common machine learning library was used for training the RNN.

5

Results


CBOW Embedding SimLex-999 Similarities
Pearson Coefficient
Spearman Coefficient
Correlation coeffeicient
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
win3-min10-size100
win3-min10-size50
win3-min5-size50
win3-min1-size50


Skip-Gram Embedding SimLex-999 Similarities
Correlation coeffeicient
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
Pearson Coefficient
Spearman Coefficient
win3-min5-neg3-size100
win3-min1-neg3-size100
win3-min5-neg3-size50
win3-min1-neg3-size50

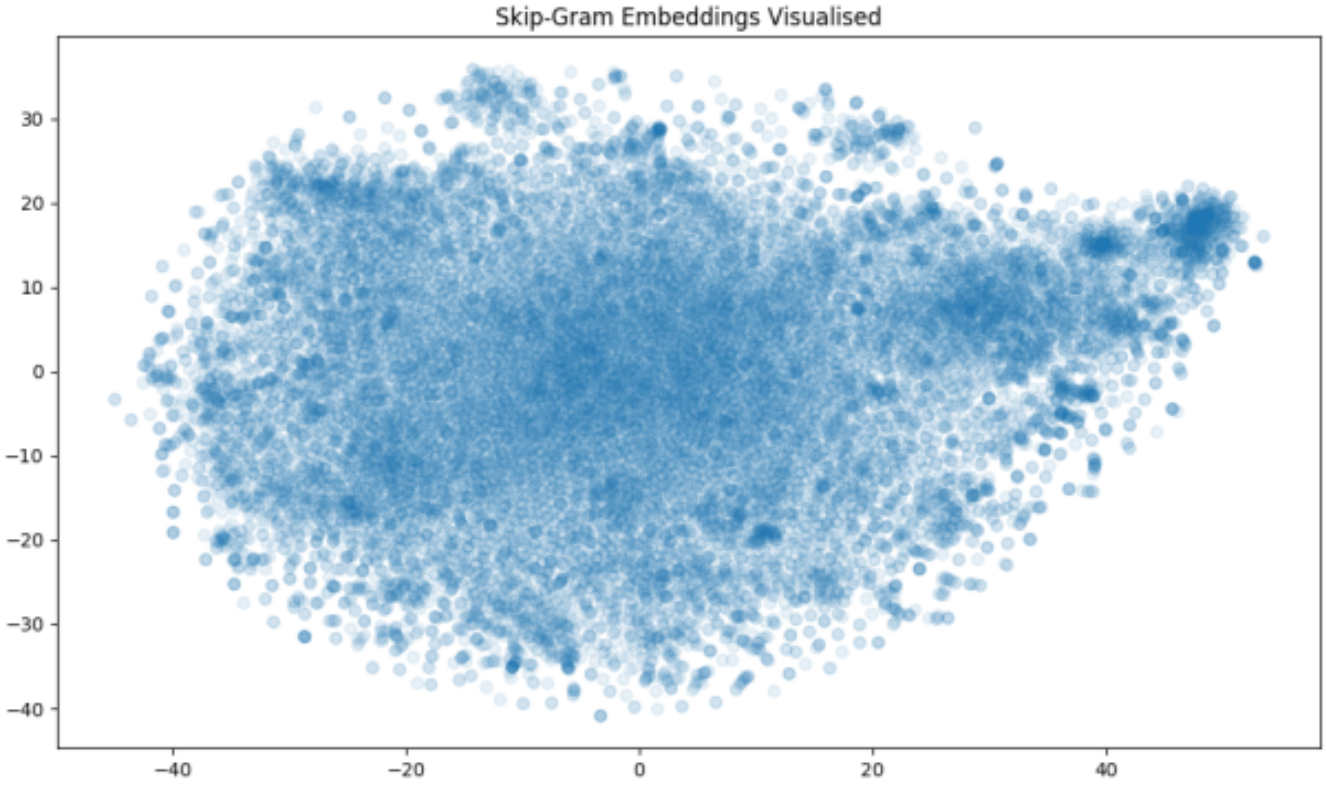


Figure: The best performing skip gram embeddings visualised using t-distributed stochastic neighbour embeddings for dimensionality reduction

5 The Best Word Embeddings

Parameters	Parameter Value	
Training Model Type	Skip-Gram	CBOW
Context Window Sizes	3	3
Minimum Word Counts	5	10
Embedding Dimension Sizes	100	100
Quantity of Negative Samples	3	—
Loss	16629545	4832437

Figure: The combination of hyperparameters that resulted in the lowest training loss values for skip-gram and continuous bag-of-words.

In the study, we showed that whilst CBOW resulted in a lower loss value, this is due to the model architecture. The SimLex-999 correlations showed in practice that Skip-Gram was better performing.

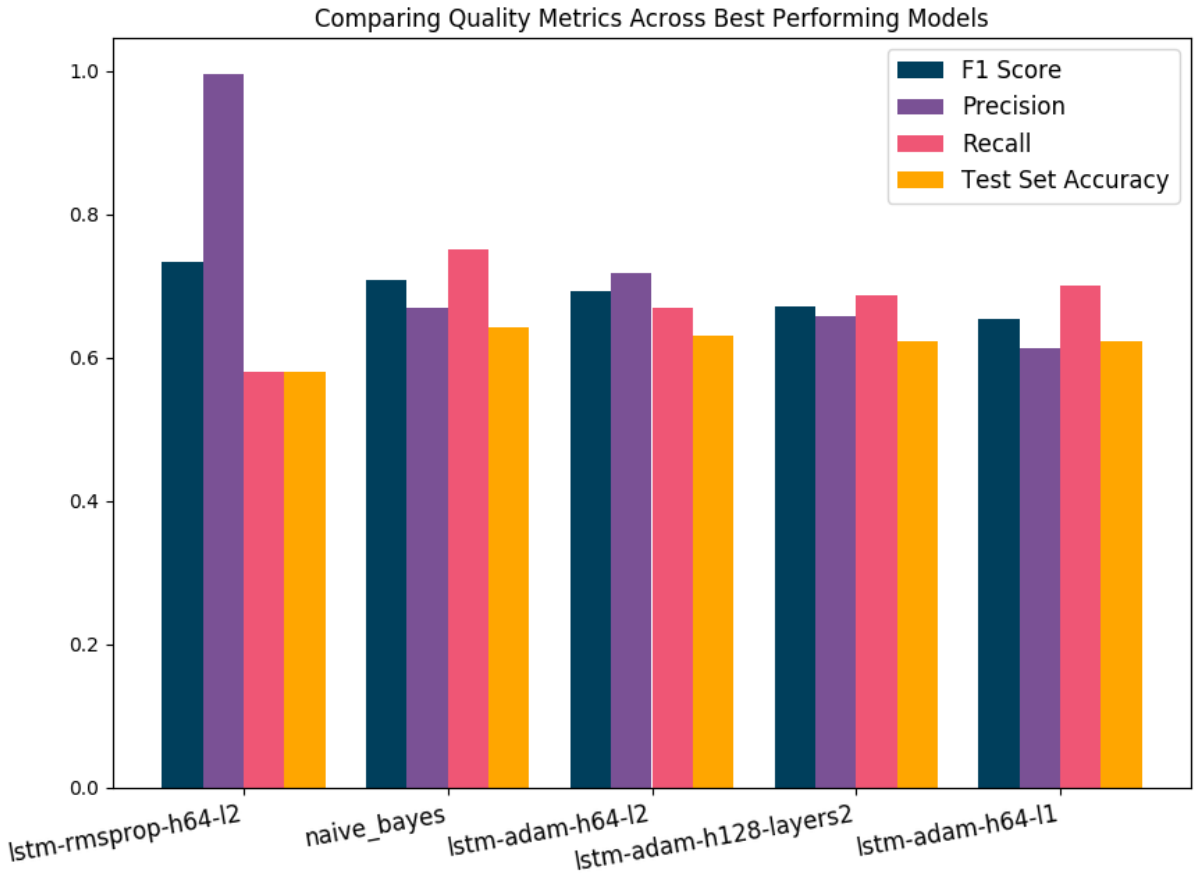


Figure: The primary performance metrics of the best performing RNN models, compared with a Naive Bayes.

Model Name	Guardian	Guardian 2019
lstm-adam-h64-l1	0.4438	0.4658
lstm-adam-h128-l2	0.4487	0.4743
lstm-adam-h64-l2	0.4742	0.4964
lstm-rmsprop-h64-l2	0.5171	0.5175
naive-bayes	0.4138	0.5289

Model Name	Daily Mail	Daily Mail 2019	BBC
lstm-adam-h64-l1	0.5827	0.5831	0.5263
lstm-adam-h128-l2	0.6072	0.6073	0.5457
lstm-adam-h64-l2	0.6054	0.6029	0.5525
lstm-rmsprop-h64-l2	0.5166	0.5171	0.5168
naive-bayes	0.8336	0.7475	0.6581

Figure: The predictions of the best performing RNN models and Naive Bayes, over certain newspapers with known biases, with 2016 and 2019 data. A prediction of 0 indicates an entirely pro-remain outlet, a prediction of 1 is entirely pro-leave.

6
Conclusions

6 Flaws in Study

- Broad application of ground truths
- Models can be inclined to learn writing styles rather than political biases
- Lack of input data can make it tricky to learn complex relationships in data.

6 Conclusions

- We are able to show that it is possible to predict political biases in certain newspaper outlets in the UK. This is the case for not only the 2016 referendum, but also on more recent data.
- We are able to show that the BBC has a slight pro-leave leaning based on our models
- We can see that simple LSTM architectures with the ADAM optimiser create the best results
- However, similarly high quality results can be found using a Naive Bayes.

6 I also wrote a blog post about this!

https://www.rory.how/blog/i-wrote-a-brexit-political-bias-classifer-and-it-sucked-heres-why/

6 References

[1] British broadcasting corporation section 4: Impartiality, 2019. [Online; accessed 9-August-2019].

[2] Circulation of newspapers in the united kingdom (uk) as of april 2019 (in 1,000 copies), 2019. [Online; accessed 9-August-2019].

[3] How left or right-wing are the uk's newspapers?, 2019. [Online; accessed 15-August-2019].

[4] CHUNG, J., GULCEHRE, C., CHO, K., AND BENGIO, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014).

6 References II

[5] Deacon, D., Wring, D., Harmer, E., Downey, J., and Stanyer, J. Hard evidence: analysis shows extent of press bias towards brexit.

[6] Hochreiter, S., and Schmidhuber, J. Long short-term memory. *Neural computation 9*, 8 (1997), 1735–1780.

[7] Wikipedia contributors. Endorsements in the 2010 united kingdom general election — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Endorsements_in_the_2010_United_Kingdom_general_election&oldid=909888211`, 2019. [Online; accessed 14-August-2019].

6 References III

[8] WIKIPEDIA CONTRIBUTORS. Endorsements in the 2015 united kingdom general election — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Endorsements_in_the_2015_United_Kingdom_general_election&oldid=909888618, 2019. [Online; accessed 14-August-2019].

[9] WIKIPEDIA CONTRIBUTORS. Endorsements in the 2017 united kingdom general election — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Endorsements_in_the_2017_United_Kingdom_general_election&oldid=909889554, 2019. [Online; accessed 14-August-2019].