Open IIT Data Analytics
IIT KHARAGPUR
2022-23

TENSORSLOW
REPORT SUBMISSION

Index

1. Introduction

The problem statement is about determining if 2 sentences are similar in terms of gender biasness. Approaches based on textual similarity have been used to solve this problem by applying various models in the field of Natural Language Processing (NLP).

2. Dataset

- The given dataset has 2000 sentences and 155951 pairs of sentences labelled according to the similarity between them based on gender biasness.
- The maximum frequency of a sentence in a pair is 501 and the minimum frequency of a sentence in the pairs is 0
- The number of sentences having 0 frequency in the training pairs is equal to 244.
- The sentences not used in the training dataset are used in the test dataset and there are 5000 pairs consisting of permutations of these 244 sentences
- The number of similar pairs were 80400 and the rest 75551 pairs were not similar in terms of gender biases in the training dataset
- As the labels are based on the similarity, it is not possible to determine the absolute labels of each sentence in terms of gender biasness.

3. Methodology

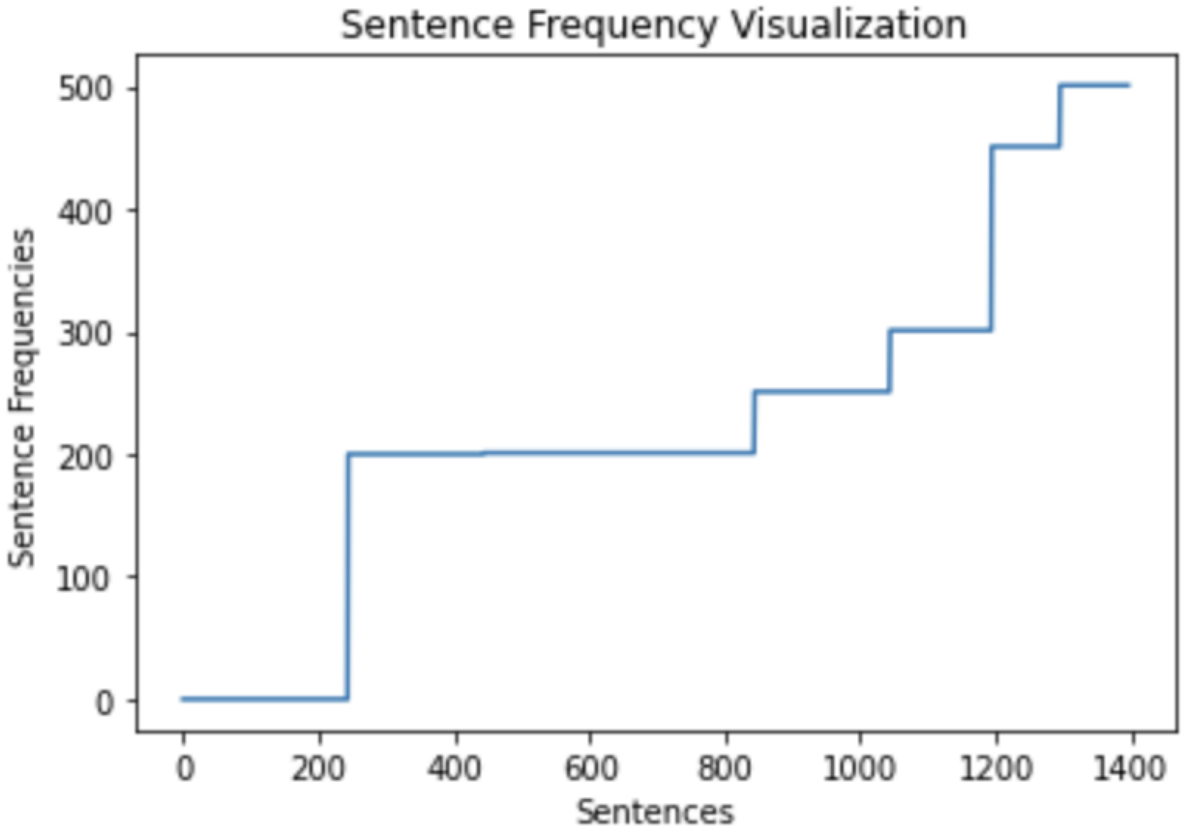
3.1 Approach

Our approach comprises two deep learning based methods and a rule based method. The first two methods employ a Siamese network and a MPNet base transformer respectively whereas on the other hand the third method is a rule based one that decides biasness based on the nouns and pronouns found in the sentence matched with a manually defined list of biased words.

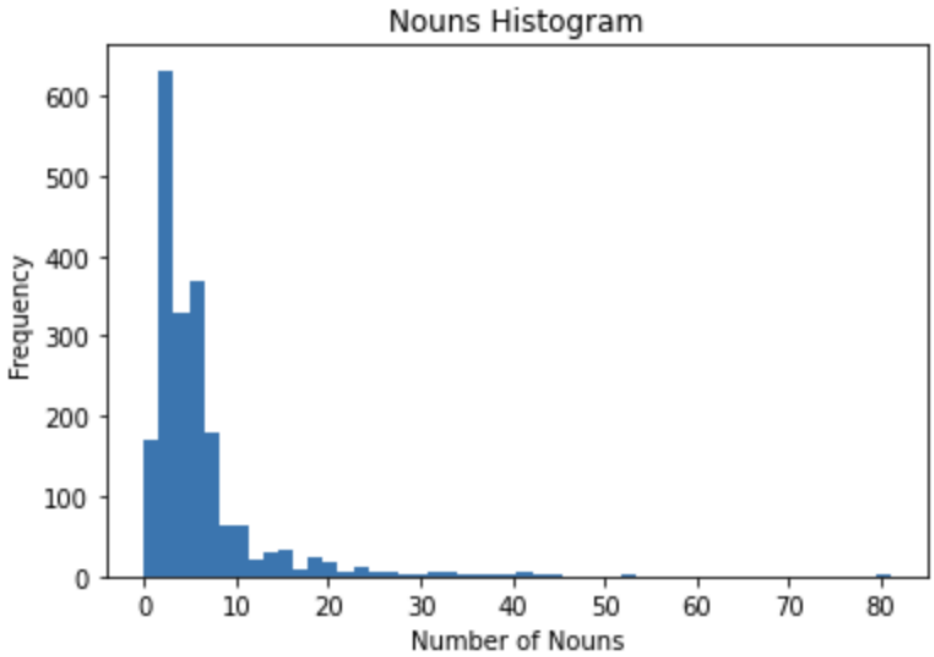
3.2 Data Pre-Processing

- The file containing texts and ids is converted into a dictionary with the sentence ids as the keys so we can easily access texts from their corresponding keys.
- The training dataset is converted into a dataframe containing the text ids and the labels of the pairs.
- The sentences corresponding to the ids are added using the earlier created dictionary to facilitate easy analysis, and the text ids are dropped from the dataframe.

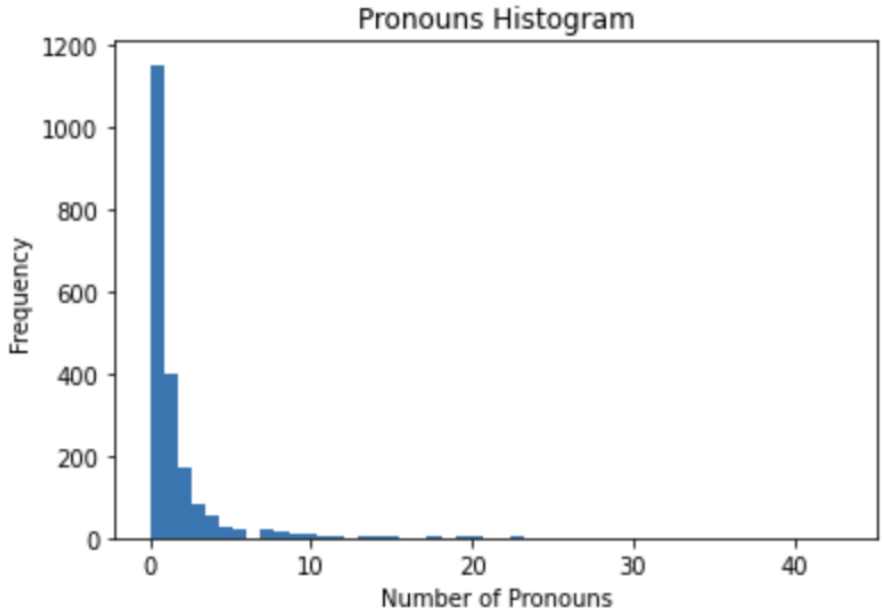
4. Data Visualization



The frequency distribution of sentences in pairs in the training dataset



This plot shows the histogram of the noun count over all the 2000 sentences. The average number of nouns per sentence is 5.775

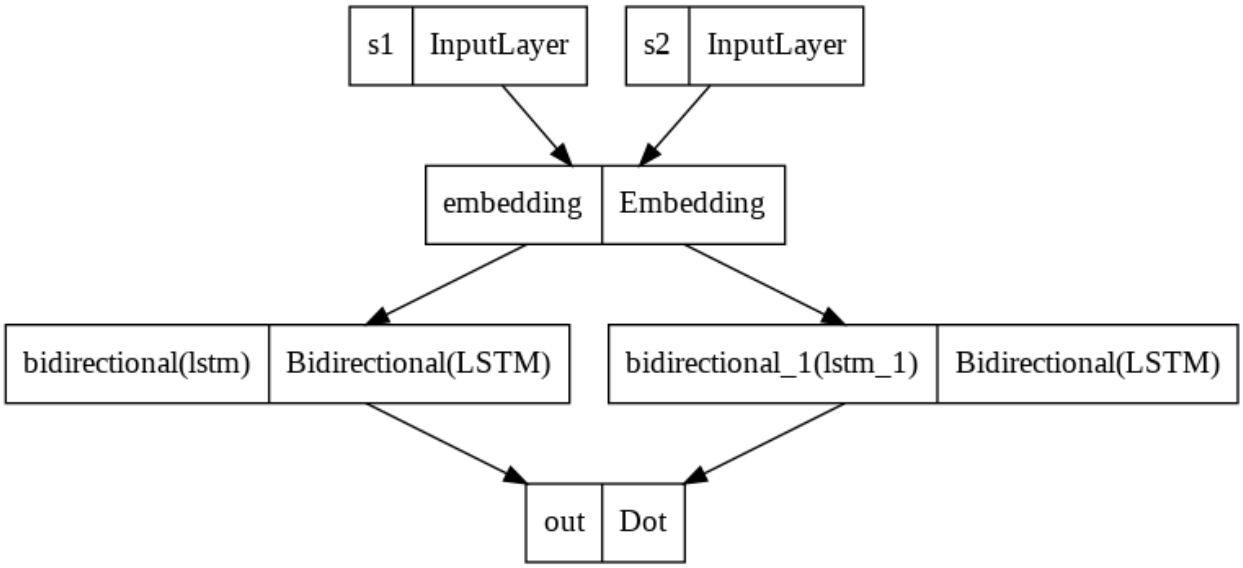


The plot represents the histogram of pronoun count over all 2000 sentences. The average number of pronouns per sentence is 1.275

5. Model & Results:

Method A:

1) The first method that we used was the Siamese network. Just as Siamese twins are connected, so are Siamese networks. Siamese networks are a special type of neural network architecture. Instead of a model learning to classify its inputs, the neural networks learn to differentiate

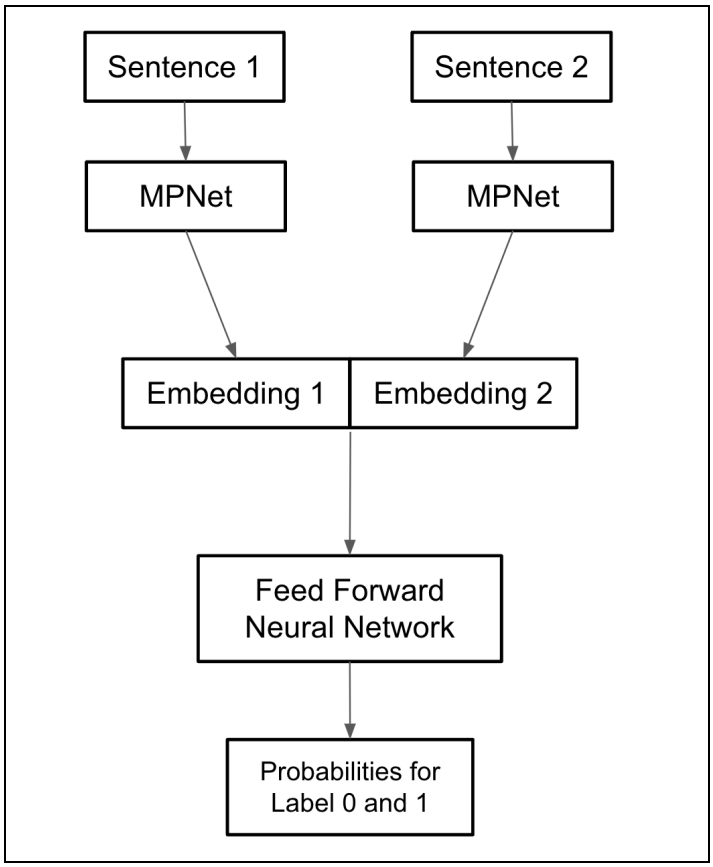


between two inputs.

2) We have used pretrained glove embeddings and we have put training of the embedding layer to false.then we have used two different bert models to create final embeddings. Then we have taken the dot product of both embeddings to generate the final output.
3) It gives a number between -1 to 1. If it is less than 0 we are predicting 0 else 1.By doing this we have increased the accuracy but still we can see scope of improvement.

Method B:

1) This is a very simple and intuitive method which uses sentence transformers namely MPNet-base.
2) A pretrained MPNet-base sentence transformer is loaded and is used to find the embeddings of all the sentences given in the corpus. These embeddings are then stored in a pkl file. Note that this operation has to be performed only once in the method pipeline.
3) The pkl file is loaded, and a simple 2 layer Feed Forward Neural network is trained for biasness prediction in our task.
4) The embeddings of the given sentence pair are concatenated and passed through the FFN which gives two output logits corresponding to the two labels of prediction 0 and 1.
5) The training is performed over 400 epochs, and it gives a training accuracy of around 98% on a test split of the training set.



Method C:

On careful observation of the dataset, it was noted that a presence of few words, especially pronouns in both sentences make both of them very similar in terms of gender biasness. Using this observation, a list of such words were used to check the gender biasness of the sentences and similarity was predicted based on it on the training dataset. This rule based approach decides based on the presence of particular pronouns present in sentences. Using this approach, an accuracy of 93.98% was obtained.

Example -1

838, 889 has been labelled as relatively unbiased(biasness=0) as both the

statements have been flagged as biased.

838 : Man is mortal. Everyone has to die some day or the other. But one must resolve to lay down one s life in enriching the noble ideals of self-respect and in bettering one s human life... Nothing is more disgraceful for a brave man than to live life devoid of self-respect.

Here, we can see that the usage of the word ‘Man’ causes the statement to be flagged as biased.

889 : India won her independence from the British .

Here, India is referred to as ‘her’, causing the statement to be flagged as biased.

Example -2

1236, 1276 has been labelled 0 as both the statements have been flagged as unbiased.

1236 : It is the duty of the home manager to take proper care of them .

We can see that the sentence is unbiased.

1276 : That is why the other name of water is life .

We can see that this sentence is also clearly unbiased

Example -3

1249, 1266 has been labelled 0 as both the statements have been flagged as unbiased.

1249 : NOTE : During the rst four counts (-) the leziem and body should swing from left to right and during the second four counts (-) .

We can see that the statement is unbiased.

1266 : Husk or bran water is also used or a cleansing material .

We can see that this sentence is also clearly unbiased

Example -4

903,1312 has been labelled 1 as the first statement is biased whereas the second statement is unbiased.

903 : . A person needs a lens of power . dioptres for correcting his distant vision. For correcting his near vision he needs a lens of power +. dioptre. What is the focal length of the lens required for correcting (i) distant vision, and (ii) near vision?

The assumption that the person in question is male causes the text to be labelled as biased.

1312 : These chemical compounds are absorbed by the bacteria and fungi as nutrients .

The text is clearly unbiased.

Example -5

936, 1342 has been labelled 1 as the first statement is biased whereas the second statement is unbiased.

936 : This is the biggest man made canal linking the Mediterranean Sea and Red Sea .

The phrase ‘man-made’ causes the text to be labelled as biased.

1342 : Bangladesh is a country of six seasons .

The text is clearly unbiased.

Example -6

938, 1305 has been labelled 1 as the first statement is biased whereas the second statement is unbiased.

938 : The victim may look as if he is asleep but his body will not execute to any movement .

The assumption that the victim in question is male causes the text to be labelled as biased.

1305 : These gases collected around the earth over millions of years and became the atmosphere .

The text is clearly unbiased.

6. Conclusion

Through the deep learning and rule based approach we tried and tested various methods and showed good results on the validation split. The transformer (MPNet) based approach gave the best possible accuracy and was used for the final CSV submission but the rule based approach also doesn't fall far behind which adds on to the explainability of the transformer based pipeline.

ANNEXURE

Word embeddings reflect human prejudices, including gender bias, according to earlier research. Studies examining the existence of certain gender bias categories in word embeddings across various domains are scarce, nonetheless.

The WEAT bias detection approach can be used on four sets of word embeddings trained on corpora from four distinct domains—news, social networking, biomedicine, and a gender-balanced corpus derived from Wikipedia. Even though this method is very promising on paper showing good results on benchmark corpuses, it couldn't be used in our problem statement as it relies on pretraining on external word corpuses.