Group members: Vikash Rajput (190216) Krishnav Garg (1901091)

SARCASM IDENTIFIER

INTRODUCTION

Sarcasm is the use of words that signify the exact opposite of what you are trying to express, usually to insult someone, or just be funny. We will use the '**News Headlines Dataset For Sarcasm Detection**' dataset to identify by building model(s) for the same

Literature Review

i. Most researchers created their own datasets since there is no standard available datasets in the domain of sarcasm identification. Context and content-based linguistic features were used in most of the studies.
ii. This review shows that n-gram and parts of speech tagging techniques were the most commonly used feature extraction techniques.
iii. However, binary representation and term frequency were utilized for feature representation whereas Chi squared and information gain were used for the feature selection scheme.
iv. Moreover, classification algorithm such as support vector machine, Naïve Bayes, random forest, maximum entropy, and decision tree algorithm were mostly applied using accuracy, precision, recall and F-measure for performance measures.

Central Idea

Here, we will attempt to build an NLP model that can identify sarcasm from normal statements using LSTM +GloVe.

EXPERIMENTAL SETUP

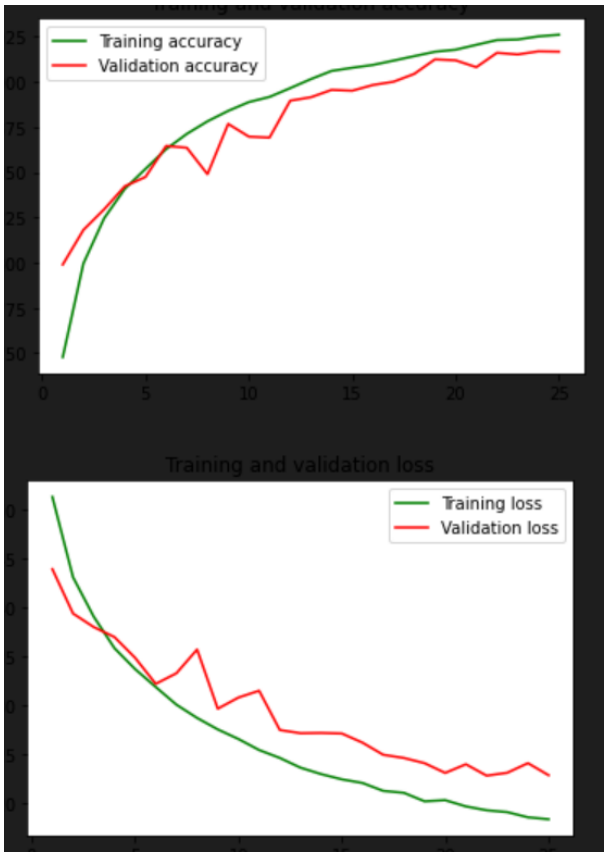
1. We clean the data by removing unnecessary data like links , emojis etc from the dataset.
2. Convert the abbreviations into full forms, adjusting punctuations and converting text into lower case.
3. Removing stop words because they don't help our ML model.
4. We convert our words into a series of vectors and pass it to the RNN which produces our prediction.

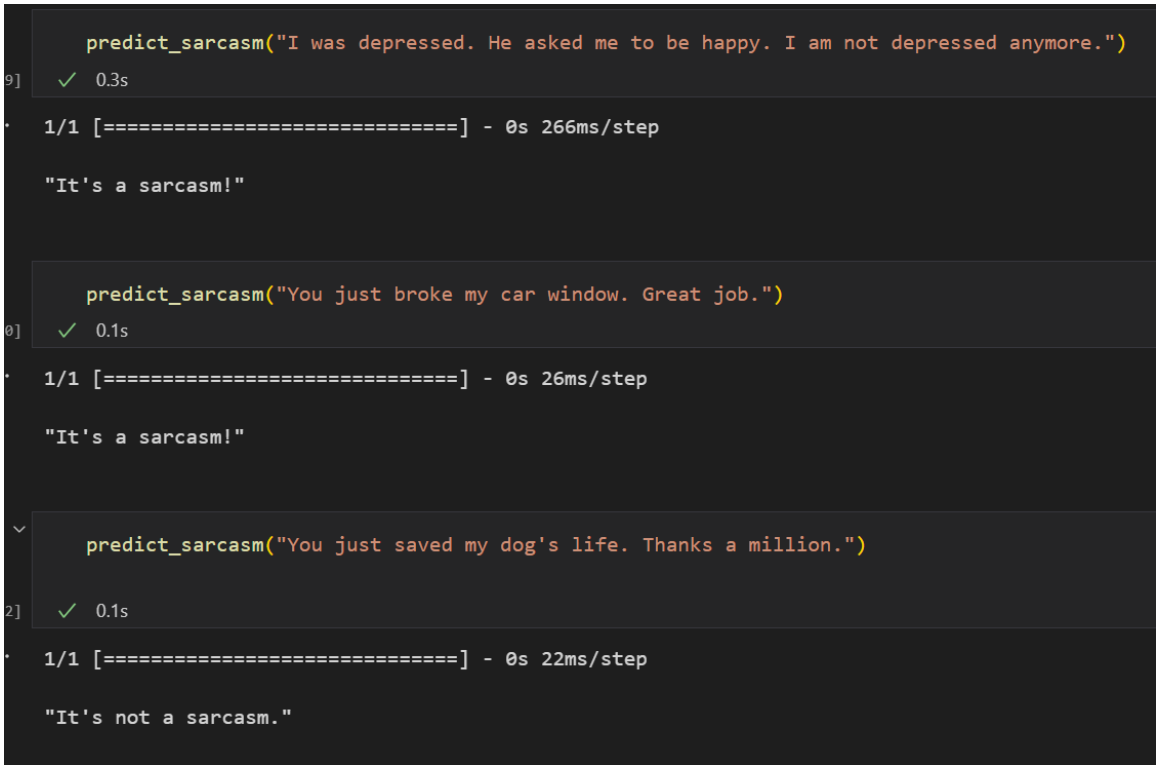
5. We can use a softmax or sigmoid function for classification purpose because we have only 2 classes.
6. We have taken 80-20 split for training and testing data.
7. We have selected first 25 words only from each new headline as it works the best .
8. Convert words into their respective vector representations and store them in a dictionary using GloVe model.
9. We build our RNN using Keras and start training it while recording its accuracy with each epoch.

Result and Conclusion

1. We visualize the training loss and validation loss graphs and conclude we can train for more epochs.
2. We also visualize the training and validation accuracy as follows:
   Training accuracy :0.9261 Validation accuracy:0.9168
3. Our models successfully detects whether the input sentence is sarcasm or not with high accuracy.

```
Epoch 25/25
1384/1384 - 33s - loss: 0.1837 - acc: 0.9261 - val_loss: 0.2286 - val_acc: 0.9168 - 33s/epoch - 24ms/step
```





References

- Boua Sarcasm identification in textual data: systematic review, research challenges and open directionszizi M, Ohtsuki TO (2016) A pattern-based approach for sarcasm detection on twitter. IEEE Access 4:5477–5488
- Ghosh D, Guo W, Muresan S (2015) Sarcastic or not: word embeddings to predict the literal or sarcastic meaning of words. Paper presented at the proceedings of the 2015 conference on empirical methods in natural language processing
- Sarcasm identification in textual data: systematic review, research challenges and open directions