Toxicity Classification Model

Our process

1. Pre trained embedding
2. Text Preprocessing
3. Bidirectional LSTM
4. Kaggle Submission

Embedding

FastText crawl 300d 2M

2 million word vectors trained on Common Crawl released by Facebook

glove.840B.300d.txt

Global Vectors for Word Representation. Pre trained word embedding model with 2.2M different words

Text Preprocessing

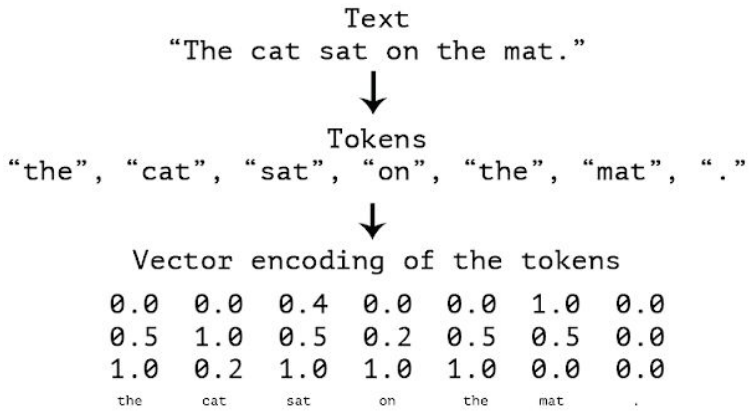
- Contraction mapping
- Swear words list
- Replace swear words list by single word ‘Fuck’.
- Lemmatizer to keep: verbs, adjectives, adverbs and nouns.

```
mapping = {"ain't": "is not",
"aren't": "are not",
"can't": "cannot",
"'cause": "because",
"could've": "could have",
"couldn't": "could not"}
```

```
swear_words = [
    ' ass ',
    ' ass-fucker ',
    ' asses ',
    ' assfucker ',
    ' assfukka ']
```

Text Preprocessing

- Keras tokenizer for train and test, excluding special characters ('!"#$%&()*+,-./:;<=>')
- Text to sequences of tokens
- Pads sequences to the same length



Bidirectional LSTMs

The idea is to split the state neurons of a regular RNN in a part that is responsible for the positive time direction (forward states) and a part for the negative time direction (backward states)

— Mike Schuster and Kuldip K. Paliwal, Bidirectional Recurrent Neural Networks, 1997

Credits

- Kaggle Notebook: *Introduction to NLP hust*. https://www.kaggle.com/duykhanh99/introduction-to-nlp-hust
- GloVe embedding: https://www.kaggle.com/takuok/glove840b300dtxt
- FastText Embedding: https://www.kaggle.com/yekenot/fasttext-crawl-300d-2m
- Mike Schuster and Kuldip K. Paliwal, *Bidirectional Recurrent Neural Networks*, 1997

Thanks!

Any questions?