AI for Fake News Detection in Social Networks

Students:

Ivancich Stefano

Marco Zanetti

Supervisor:

Prof. Pini Maria Silvia

DEPARTMENT OF INFORMATION ENGINEERING

UNIVERSITY OF PADOVA

12 May 2021

Outline

- The Problem
- The Dataset and pre-processing
- Our models
- Results
- Conclusions

The Problem - 1

Definition: Fake news is false or misleading information presented as news. It often has the aim of damaging the reputation of a person or entity, or making money through advertising revenue.

Deep Psychological base:

- "Most of the decisions that people take are not made with logic, but are made with emotions. Then the logic is used to justify the irrational choice that the individual has made."
- Fake news **leverage emotions** to manipulate people beliefs.

It's not a recent year's phenomenon:

- There are proof that they are being broadly used especially in politics at least since elections of Cicerone in 64 b.c. (Commentariolum Petitionis)
- Probably they are going to be used in the future.

The Problem - 2

Important facts about fake news:

- Fake News on Social Networks spread rapidly, with a peak in the first **2 hours**.
- There is a strong economical and **political interest** in their use. (E.g., Cambridge Analytica, Trump elections, Russian Ads,…)
- **Manual techniques** such as debunking and fact checking websites are very **slow**, not applicable on a large scale, and can be easily blocked.
- Social Network bubbles make people isolate themselves in their beliefs and prejudices that can be easily manipulated.
- The average person doesn't have the time, want, resources or intellectual abilities to verify the sources and study deeply a topic, they just take each statement they read as true.

highlight the **need for an automatic system** for detecting Fake News and hoaxes on Social Networks platforms.

The Dataset

We've worked with the Covid19 Fake News dataset from [Patwa et al., 2021], which contains tweets about the pandemic labelled either Real or Fake.

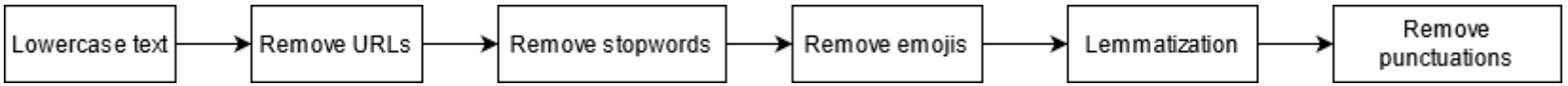
The dataset has been divided as follows:

- A training set, used to train our models, composed of 6420 tweets;
- A validation set, used to choose hyperparameters and validate our models, composed of 2140 tweets;
- A test set, used to assess the performance of our models, composed of 2140 tweets.

The dataset is balanced, that is, the number of Fake news and True news it's almost even. This is a desirable property for classification task as it makes harder for the model to solely rely on statistical properties of the data for prediction.

Data pre-processing

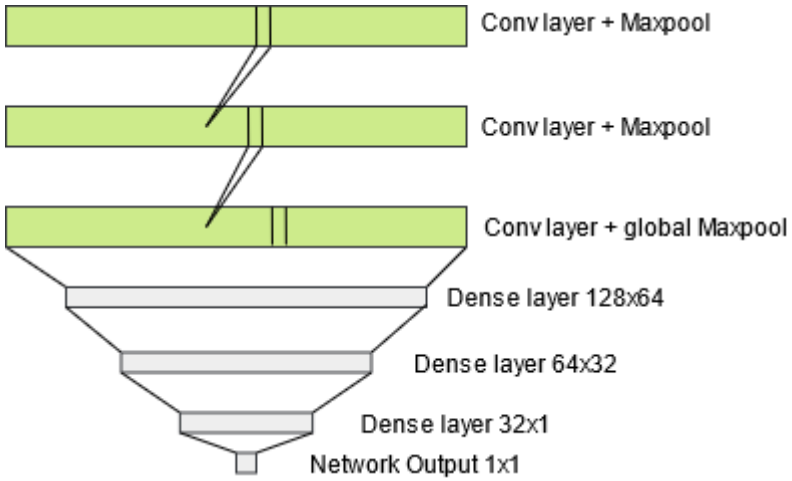
We follow standard Natural Language Processing procedures to make the text easier to process for our models.



- We convert the whole text to lowercase (e.g 'A' becomes 'a');
- We remove all URL;
- We remove all stopwords (e.g 'and', 'or', 'the');
- We remove all emojis;
- We lemmatize the text (e.g 'am', 'are', 'is' become 'be');
- We remove punctuations (e.g '.', '#', '!'),

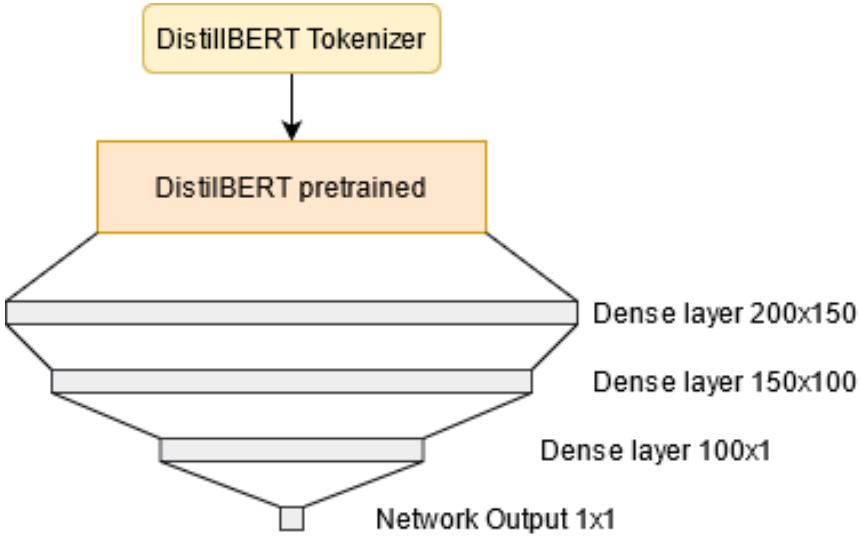
1D CNN

One-dimensional Convolutional Neural Network have been used widely in Natural Language Processing [Yoon Kim, 2014]. Text is tokenized trough an embedding layer which uses pretrained word embeddings. We deploy a network composed of 3 convolutional layers (followed by Max pooling) with kernel size of 5 and 3 Dense layers (with dropout), using adam optimizer and binary cross entropy loss function.



DistilBERT

We experimented with HuggingFace's implementation of the attention model DistilBERT [Sanh et al., 2019]. DistilBERT tokenizer encode the data so that the model can process it. We imported the pretrained weights for DistilBERT and stacked 3 Dense layers with dropout on top of the model to fine-tune for our classification task, using adam optimizer and Binary Focal loss.



Results

Experiment results shows that both models largely outperform the baseline. DistilBert achieve greater accuracy at the cost of longer training and execution time and the needs of an advanced hardware (DistilBERT can take many hours to train without using a GPU).

	Accuracy %	**# Parameters**	**Speed (s)**
Baseline (random guess)	50	/	/
1DCNN	87.8	4,804,087	0.87
DistilBERT	94.7	66,363,649	21.67

Conclusions

Conclusions

- Our models show that they can learn effectively by surpassing by a large margin baseline results;
- The attention model largely outperform the simpler 1DCNN model, as is to be expected, but 1DCNN is extremely faster and less computational extensive to train.

Future Work

- Try different hyperparameters during training that may lead to better performance;
- Implement different architectures (e.g recurrent networks);
- Deploy the model trough a high-level platform to allow user to experiment with it.