Project Report

Twitter Sarcasm Classification Challenge

Sahil Rishi - sahilr2@illinois.edu

Work Done:

I have made 11 submissions for my challenge (username: reckoner) and have achieved a best rank of 8.

After sometime I started investigating models with lower parameters which still allow me to beat the baseline.
The current leaderboard results are with a distilbert model, Epoch-8, 'max_seq_length': 256, lr: 3e-5.

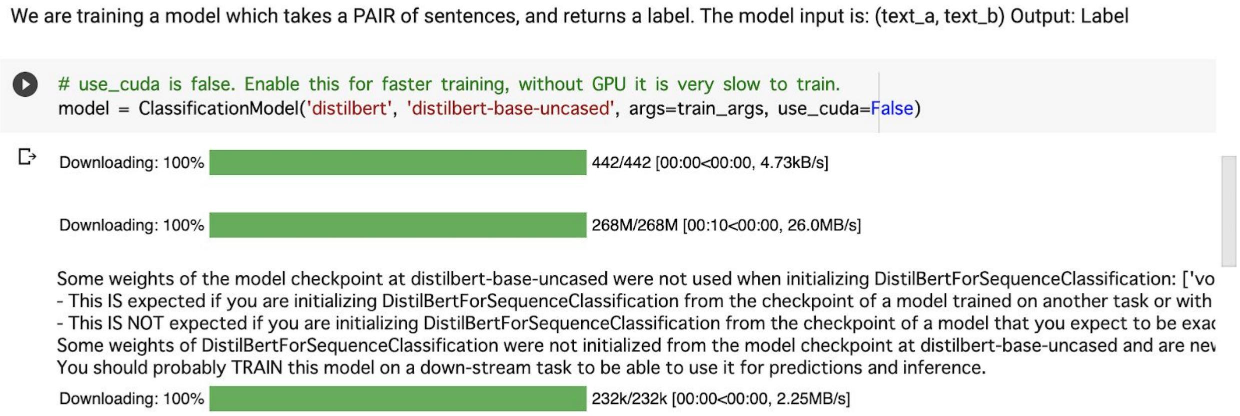
That model is also loaded into **`demo.ipnyb`** and can be loaded and run.

Task	Status	Comments/Challenge
Dataset Preparation	Done	We have to create dataset in a format so that we can try a variety of problem formulations
Framework	Done	**Model Supported:** Bert, Distilbert and RoBerta Models. **Task Supported:** Binary Classification Sentence Pair Classification
Method 1: (distilbert-base-uncased, Only Response, no pre-processing)	Done	**Got Rank 20** and F1_Score:0.73
Method 2: (distilbert-base-uncased,Response+Context, no pre-processing)	Done	**Got Rank 8** and F1_Score:0.756
Hyper Parameter Searching	Done	Used wandb hyper-parameter tuning on lr. Sequence size manually tested.

Model

We used a transfer learning package called SimpleTransformers. This package supports creating and training huggingface transformer models and provides necessary abstractions to make the process faster.

We make use of https://simpletransformers.ai/docs/sentence-pair-classification/ module of the library.



We are training a model which takes a PAIR of sentences, and returns a label. The model input is: (text_a, text_b) Output: Label

```
# use_cuda is false. Enable this for faster training, without GPU it is very slow to train.
model = ClassificationModel('distilbert', 'distilbert-base-uncased', args=train_args, use_cuda=False)
```

```
Downloading: 100% 442/442 [00:00<00:00, 4.73kB/s]

Downloading: 100% 268M/268M [00:10<00:00, 26.0MB/s]

Some weights of the model checkpoint at distilbert-base-uncased were not used when initializing DistilBertForSequenceClassification: ['vo
- This IS expected if you are initializing DistilBertForSequenceClassification from the checkpoint of a model trained on another task or with
- This IS NOT expected if you are initializing DistilBertForSequenceClassification from the checkpoint of a model that you expect to be exac
Some weights of DistilBertForSequenceClassification were not initialized from the model checkpoint at distilbert-base-uncased and are nev
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
Downloading: 100% 232k/232k [00:00<00:00, 2.25MB/s]
```

This module trains a transformer model to predict over a pair of sentences.
The idea is to use the response of the tweet and the context as the pair of sentences.
i.e.
text_a , text_b   =>   Sarcasm/Not Sarcasm

Here text_a is **response**
text_b is **concatenation of (context 2 and context 1)**

For testing parameters I used a 80:20 split. Final training was done on all the data points.

Other Methods Tried:

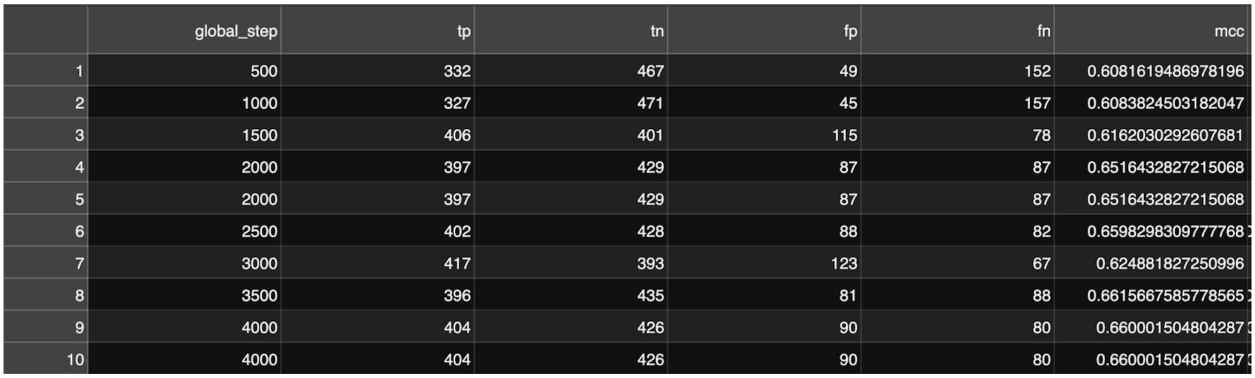
- I also tried a simple classification model(within simpletransformers library) with only the response. It gave me **F1_Score:0.73**
- I also tried **Roberta Large model**. This model did not give me a successful result. The reason for this was that as the model was very large, only small sequence lengths were fitting in the GPU (seq: 32). This proved to be too small to capture the sentence embeddings and this variation of the model failed.

- I also tried **Bert and Bert Large** model. They gave similar performance to distilbert models so I investigated only distilbert.

HyperParameter Tuning:

SimpleTransformers library provides hyperparameter tuning support with the wandb. I investigated optimal lr.
For sequence length I investigated by hand as I observed that small changes in sequence length did not affect the scores by a lot.

Optimal lr: 3.1134e-5
Sequence: 256
Epoch: 8
After 4000 steps (or 8 epochs) the model stopped automatically as we put the early stop parameter. With this when the model stops learning the training procedure stops itself.



	global_step	tp	tn	fp	fn	mcc
1	500	332	467	49	152	0.6081619486978196
2	1000	327	471	45	157	0.6083824503182047
3	1500	406	401	115	78	0.6162030292607681
4	2000	397	429	87	87	0.6516432827215068
5	2000	397	429	87	87	0.6516432827215068
6	2500	402	428	88	82	0.6598298309777768
7	3000	417	393	123	67	0.624881827250996
8	3500	396	435	81	88	0.6615667585778565
9	4000	404	426	90	80	0.660001504804287
10	4000	404	426	90	80	0.660001504804287

(Results reported on 1000 samples withheld from the training data)

Due to GPU costs I only performed hyperparameter tuning on sentence pair classification distilbert model.

Summary

As I wanted to use large transformers models, I made use of a specialised library which abstracts many functions required for transfer learning. Due to this the task of using complex models such as distilbert, bert, Roberta become really easy and straightforward.