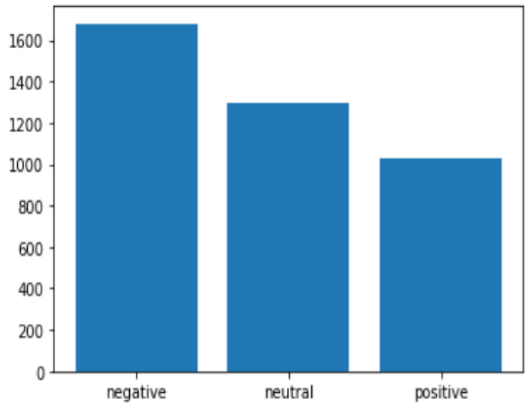
Training Methodologies

Data Analysis

Few Observations:

1. There were no Nan values in the train dataset
2. The dataset is minorly imbalanced between negative and positive sentiments



Experiments

Important points of the paper Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence

Main idea is to construct an auxiliary sentence from the aspect and convert it to a sentence-pair classification task. Fine tuned using BERT (can use different types)

Introduction

The task is to divide the process into 2 steps:

1. Determine the aspects associated with the target
2. Resolve the polarity to a given target

Problem Description

Targeted Aspect Based Sentiment Analysis (TABSA)
A sentence consists of a series of words and some of the words are pre-identified targets and we try to identify the sentiments associated with them.

I am trying to experiment with varying amounts of aspects and adapting my data to fit the current algorithm.

My Results

I am using two metrics,

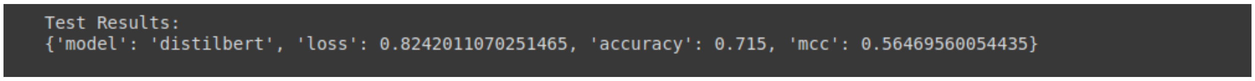
1. Accuracy
2. Mathew's Correlation Coefficient

$$\mathrm{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Why? Majorly because it contains more information than F1-score. More the MCC, better the model

Epoch Number	Accuracy (train)	MCC (train)
0	0.678	0.509
1	0.717	0.563
2	0.726	0.726
3	0.711	0.556

Results on the first approach (Took around 3 hours to train)



```
Test Results:
{'model': 'distilbert', 'loss': 0.8242011070251465, 'accuracy': 0.715, 'mcc': 0.56469560054435}
```

Evaluation took around 6 minutes

The reason for such huge time usage is majorly because there are a lot of different aspects and the above approach is tailored towards a fixed number of aspects.

Vanilla BERT and Sentence Pair Classification

Approach

BERT-SPC feeds sequence “[CLS] + context + [SEP] + aspect + [SEP]” into the basic BERT model for sentence pair classification task.

Results

The metrics used here were Accuracy and macro F1.

Epoch number	Val accuracy	Macro F1 (Val)
0	0.7113	0.6918
1	0.7262	0.7250
2	0.7225	0.7206
3	0.7438	0.7384
4	0.7450	0.7361

The above approach took significantly less time to train and the above approach also had worse performance during training as compared to this method. So I decided to go on with this and proceed ahead with the inference task.