CS772: DL4NLP Assignment Evaluation

Team Id: 48
Manoj Bhadu, 170010036
Piyush Onkar, 160050012

Date: 12/05/2021

Problem Definition

1. Sentiment Analysis using Transformers

2. Compare the performance of Transformer Model with the models created in the last assignment.

3. Error analysis of the Transformer Model.

System Architecture

```
Model: "model"
__________________________________________________________________________________________________
Layer (type)                    Output Shape         Param #     Connected to
==================================================================================================
input_ids (InputLayer)          [(None, 512)]        0
__________________________________________________________________________________________________
attention_mask (InputLayer)     [(None, 512)]        0
__________________________________________________________________________________________________
bert (TFBertMainLayer)          TFBaseModelOutputWit 109482240   input_ids[0][0]
                                                                 attention_mask[0][0]
__________________________________________________________________________________________________
lstm (LSTM)                     (None, 512, 32)      102528      bert[0][0]
__________________________________________________________________________________________________
lstm_1 (LSTM)                   (None, 16)           3136        lstm[0][0]
__________________________________________________________________________________________________
outputs (Dense)                 (None, 5)            85          lstm_1[0][0]
==================================================================================================
Total params: 109,587,989
Trainable params: 105,749
Non-trainable params: 109,482,240
__________________________________________________________________________________________________
```

Libraries Used for this Assignment

- Pandas
- Numpy
- Tensorflow
- Sklearn
- Seaborn
- Matplotlib
- Tkinter
- Transformers

Results

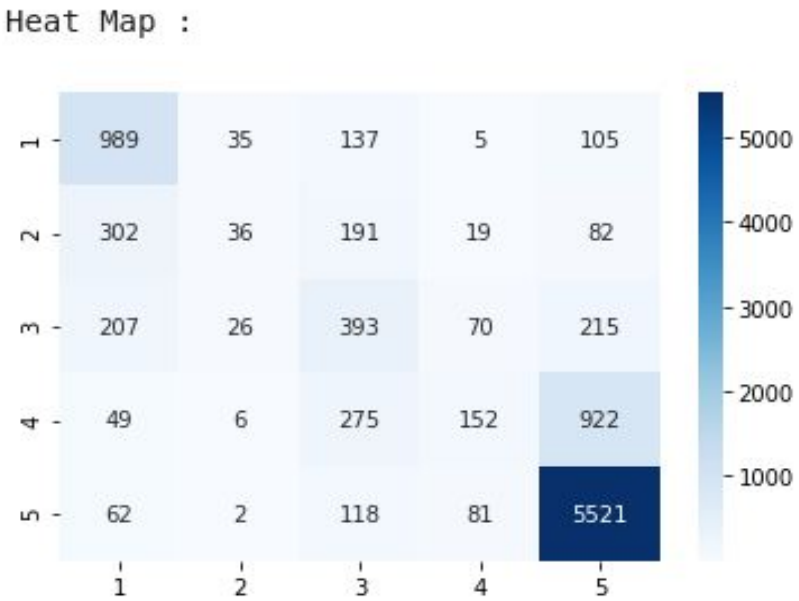
	Models					
Metrics	RNN	LSTM	Bi-LSTM	GRU	Bi-GRU	Transformer
Precision	0.4	0.59	0.6	0.6	0.59	0.66
Recall	0.58	0.66	0.66	0.66	0.65	0.71
F1-Score	0.44	0.6	0.6	0.61	0.6	0.66
Accuracy	0.58	0.66	0.66	0.66	0.65	0.71

**Pre-trained Word Embeddings are used for models RNN, LSTM, Bi-LSTM, GRU, Bi-GRU*

Qualitative Analysis with some examples

Sentence	Predicted	Real
Caution! Do not wash this with anything else (including dark towels) because the black will bleed over everything. Went right in the trash.	1	1
Very big, plenty buttons for the neck, wraps around the back of the shoulder enough for me.	5	5
Keeps the hair off good, but the hair seems to stick to it after the hair cut in spite of vigorous shaking.	3	3
I really love the headbands they were just a little thin.	4	4
Turned my finger green... I coated it in clear nail polish and it still turned my finger green..	1	1
It broke in half the second day. =/ It was cute though.	1	2
Not great. It gives me a headache... probably because it's bee venom. Plumps just a little.	3	2
Arrived earlier than expected and looks great. Should suit my needs perfectly.	5	4

Confusion Matrix for Transformer Model



Confusion Matrix Analysis

- The accuracy of predicting class label ‘2’ is very less because the training dataset has less examples for label ‘2’. Also the model is trained for 2 epochs. The training time for each epoch is more than 2 hours on GPU.
- The class label ‘2’ most of the times would predict either ‘3’ or ‘1’ depending upon the number of positive and negative words in the sentence.
- For each label, more than 80% of examples lie in the range of (label_value + 1) and (label_value - 1) except for class ‘3’ label.

Demo and Lime Analysis

- Arrived earlier than expected and looks great. Should suit my needs perfectly.
  The Lime Analysis predicted “5” which is equal to the label predicted by Transformer model. The true label is “4”.

- Not great. It gives me a headache... probably because it's bee venom. Plumps just a little.
  The Lime Analysis predicted “3” which is equal to the label predicted by Transformer model. The true label is “2”.

- It broke in half the second day . =/ It was cute though .
  The Lime Analysis predicted “4” which is not equal to the label “1” predicted by Transformer model. The true label is “2”.
  The LIME analysis predicts “4” because of many positive words captured by the LIME.