Real time Emotion Classification on portable devices

Kartavya Bhatt

December 2021

Abstract

AI systems these days are not very successful in understanding human emotion in real-time. It leads to negative implications on various systems like virtual assistants and chatbots as well. The sophisticated AI systems are good at emotion recognition but requires high computation power which is not an issue for cloud based applications but when it comes to portable devices like mobile phones, or embedded systems for that matter there is a lack of better performing AI. We try to address the problem by using BERT model and attaching a classifier neural network with just 2 layers. EmotionLines corpus is used to train, validate and test the architecture. We understood the issues with the dataset and solved the issues hence attaining 70% accuracy with some constraints in classification labels. For the future work, such systems can help in learning to understand the user emotions and associate them with virtual assistant errors.

1 Introduction

Due to the contextual nature of emotion in the utterance, emotion classification is difficult. There are many words that can represent different emotions depending upon the context e.g. 'What', 'How'. We choose to use EmotionLines corpus to incorporate context but as we are addressing the real time classification problem we only use prior context. We hypothesize that recent advanced NLP models like BERT can effectively incorporate prior context.

2 Dataset

We used EmotionLines dataset[1], which consists of dialogues from Friends show and each dialogue is labeled. There are total 14,503 utterances with average length of 10.67 words. There are 8 labels namely: Neutral, Non-neutral, Joy, Surprise, Anger, Sadness, Disgust, Fear. The distribution of the dataset as per the labels can be seen in the Figure 1.

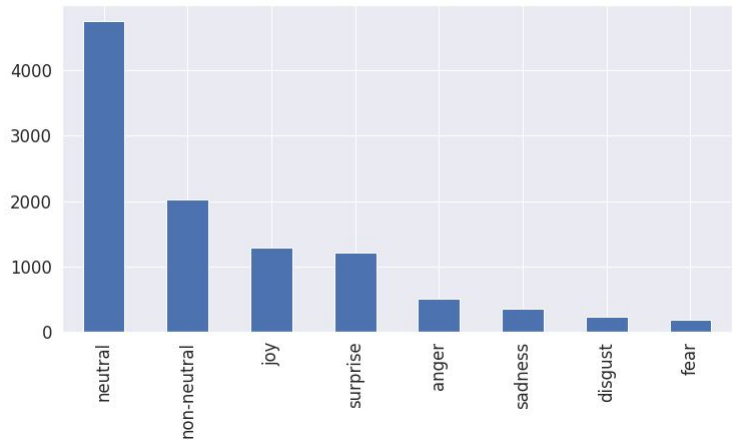


Figure 1: Class Distribution

3 Method

We have fine-tuned the BERT model and trained to make utterance level classification. The training set is defined as datapoints (u_i, e_i), where u_i is i^{th} utterance and e_i is its corresponding emotion label. Each utterance is tokenized as per BERT model's requirement. [CLS] token is added at the starting of each utterance and the tokens are mapped to ids according to BERT's vocabulary. [SEP] is appended at the end of each utterance. The encoder output corresponding to the [CLS] token is taken as an input to the classifier. i.e. Figure 2

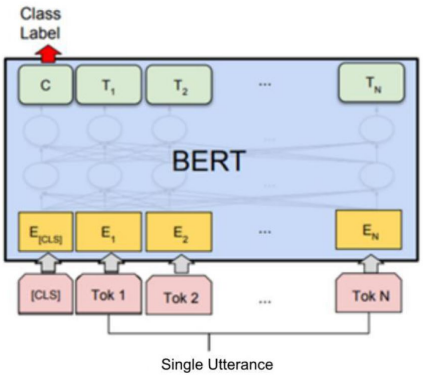


Figure 2: BERT + classifier architecture

4 Results and Discussion

We trained the model using various subsets of the dataset. As seen in figure 1, the dataset is very imbalanced. So we tried to trim the dataset to keep the classes in balance and in another experiment we removed Neutral and Non-neutral classes from the dataset. For generating the trimmed dataset we reduced the number of utterances in Neutral and non-neutral class to 500(as seen in Figure 3). When trained on the entire dataset we achieved 59% accuracy and 40% weighted accuracy. When trained on trimmed dataset model achieved 52% accuracy and 46% weighted accuracy. And for the reduced class dataset the model achieved 70% accuracy and 51% weighted accuracy.

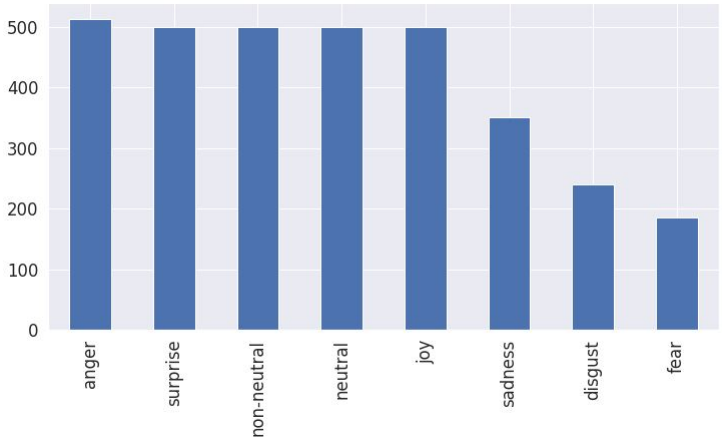


Figure 3: Class Distribution after trimming the data

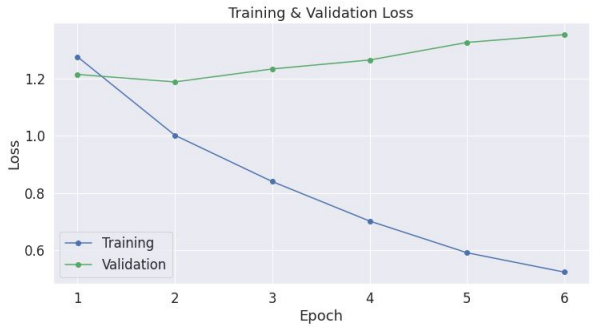
In the first experiment, when the whole dataset is used the model tends to overfit.As seen in the training loss curve the model learns till 2 epochs and then starts overfitting as the validation loss starts rising(as seen in figure 4a). This also supports the hypothesis for BERT model which says that the fine-tuning should only be done for 2-4 epochs. A similar case can be observed when trained on the trimmed dataset and removed class dataset as seen in figure 4b and 4c.

As seen in the figure 5a, the confusion matrix shows that when trained on the whole dataset the model tends to classify a utterances to neutral class. This is because of the fact that the dataset is imbalanced to neutral class. The same trend continues for the trimmed dataset.

5 Conclusion

References

[1] Chao-Chun Hsu et al. "EmotionLines: An Emotion Corpus of Multi-Party Conversations". In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association (ELRA), May 2018. URL: https://aclanthology.org/L18-1252.

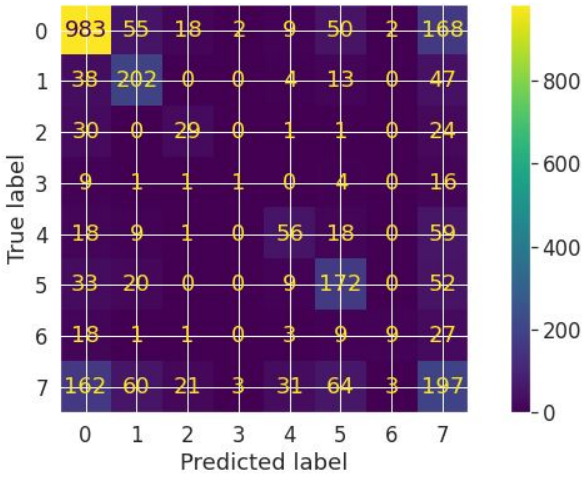


(a) Training loss curve when using whole dataset

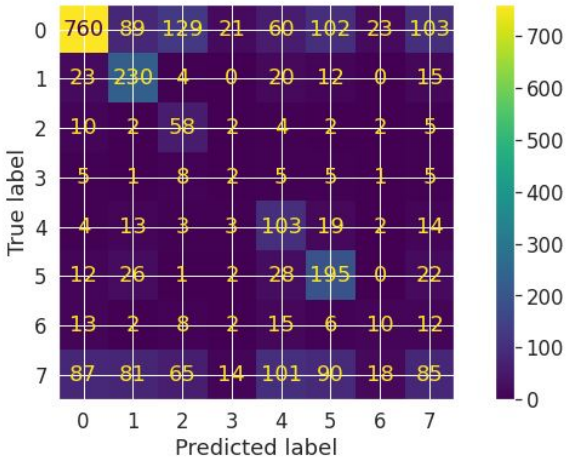
(b) Training loss curve when using trimmed dataset

(c) Training loss curve when using dataset without Neutral and non-neutral class

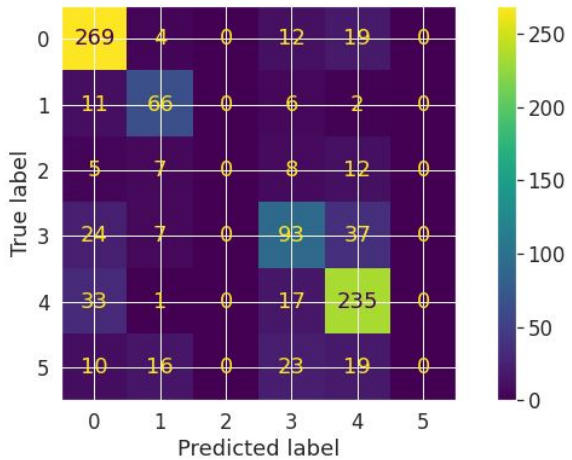
Figure 4: Training curves



(a) Confusion Matrix when model is trained on whole dataset.



(b) Confusion Matrix when model is trained on trimmed dataset



(c) Confusion Matrix when model is trained on dataset without Neutral and non-neutral class

Figure 5: Confusion Matrices