

Speech Emotion Recognition using Semantic Information

Navya (210657)

1 Implementation Details

All the codes given in the original github link by the author were ran other than evaluation. All the training codes(4 files) contained code which had attributes of tensorflow 1.x version and not compatible with tensorflow 2.x version which had to be converted and to work with tensorflow 2.x version and python 3.9

For new implementation, a new attention layer was added to the model after the LSTM output of the fused paralinguistic and semantic features.

This helps to decrease the processing time and also decrease the model size which reduces the computational power required by the user.

It also makes the output and training more focused and enhanced on specific and important features.

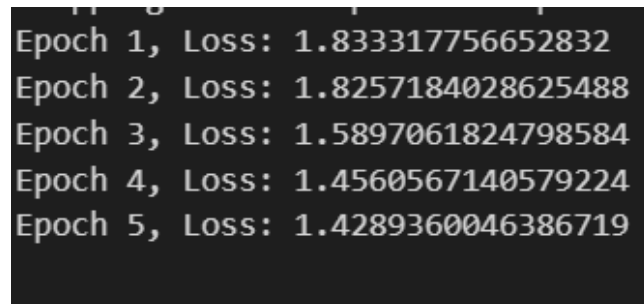
2 Dataset

Dataset provided only had the audio file and transcript file of the participants and did not contain the labels/annotations for emotional(arousal, valence, and liking) data.

The data was generated by me by studying the required format and the necessary fields. The folders(turns and labels) are both created by me, and the labels data contains all zeroes as this was to check whether the code would run.

As specified, I tried it for the segment of different emotions, but those did not contain valid transcript which was the same problem as before and thus could not be used clearly. Just using their valence and liking data did not change much, and it had the same results as the emotional data was random as compared to the transcript.

3 Results



```
Epoch 1, Loss: 1.833317756652832
Epoch 2, Loss: 1.8257184028625488
Epoch 3, Loss: 1.5897061824798584
Epoch 4, Loss: 1.4560567140579224
Epoch 5, Loss: 1.4289360046386719
```

Figure 1: Training values