Natural Language Processing 2024
Assignment-4 Report

Group No. 41

TASK 1: ERC Task [Emotion Recognition in Conversation]

DATA FORMAT

The following is an example of one dialogue. There are about 6k such examples in the training set. Firstly the emotion and utterance pairs in a dialogue can be separated.

"episode" : "utterance3492",

"speakers" : ["Phoebe","Eric","Phoebe","Eric ","Phoebe"],

"emotions" : ["surprise","fear","surprise","sadness","disgust"],

"utterances" : ["You-you u0085you had sex with Ursula?!", "Uh, a little bit. She-she-she walked in and I thought she was you and I kissed her and", "You didn't notice she was wearing different clothes?!", "Well I was just so excited to see you.", "Oh. Ew! Ew! Ew! Ugh! Y'know what? This is too weird."],

"triggers":[1.0,1.0,0.0,0.0,0.0]

U1: "You-you 0085you had sex with Ursula?!"
U2: "Uh, a little bit. She-she-she walked in and I thought she was you and I kissed her and"
U3: "You didn't notice she was wearing different clothes?!"
U4: "Well I was just so excited to see you."
U5: "Oh. Ew! Ew! Ew! Ugh! Y'know what? This is too weird."

E1: surprise
E2: fear
E3: surprise
E4: sadness
E5: disgust

MODEL 1 (Intuition and Architecture)

For each utterance, we defined a window of size max(utterances before the current utterance, 5) ending on the current utterance. This new window utterance along with the corresponding emotions was taken as the modified input. This was done to predict the emotion of the current utterance we may need to take the context of the previous few utterances that were said, as they might contain some useful information that was responsible for the emotion expressed in the current utterance.

MODEL 1 (Results)

M1 - Average Test Loss: 0.0017, Test accuracy: 0.9908, Test F1 score: 0.9908

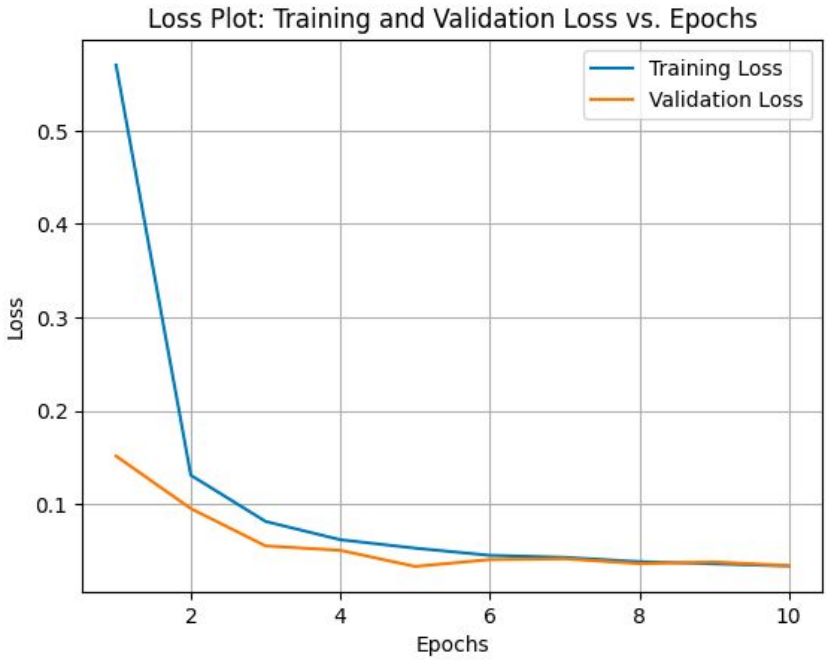

Figure 1: Model 1: Training Loss and Validation Loss

MODIFIED DATA FORMAT

U1: "You-you 0085you had sex with Ursula?!"

U2: "You-you 0085you had sex with Ursula?! Uh, a little bit. She-she-she walked in and I thought she was you and I kissed her and"

U3: "You-you 0085you had sex with Ursula?! Uh, a little bit. She-she-she walked in and I thought she was you and I kissed her and. You didn't notice she was wearing different clothes?!"

U4: "You-you 0085you had sex with Ursula?! Uh, a little bit. She-she-she walked in and I thought she was you and I kissed her and. You didn't notice she was wearing different clothes?! Well I was just so excited to see you."

U5: "You-you 0085you had sex with Ursula?! Uh, a little bit. She-she-she walked in and I thought she was you and I kissed her and. You didn't notice she was wearing different clothes?! Well I was just so excited to see you. Oh. Ew! Ew! Ew! Ugh! Y'know what? This is too weird."

E1: surprise
E2: fear
E3: surprise
E4: sadness
E5: disgust

Note: if there was a U6 then that would have the concatenation of U2 all the way up to U6

Now for the modified utterances, BERT embeddings were computed, and BERT. For Sequence Classification head was used to predict the emotion of the current utterance. The modified utterances from each dialogue were taken as the training data and the model was fine-tuned on this dataset.

MODEL 2 (Intuition and Architecture)

In this model, for each utterance, we compute the FastText representation of each word (embedding dimension=300) in the utterance. Then to compute the utterance representation we take the mean of all words in the utterance. This utterance embedding for each utterance in the data (across dialogues) is passed in a GRU model. And finally softmax is applied on the last time step output of GRU to predict the emotion of the current utterance.

MODEL 2 (Results)

M2 - Average Test Loss: 0.0367, Test accuracy: 0.5767, Test F1 score: 0.5449

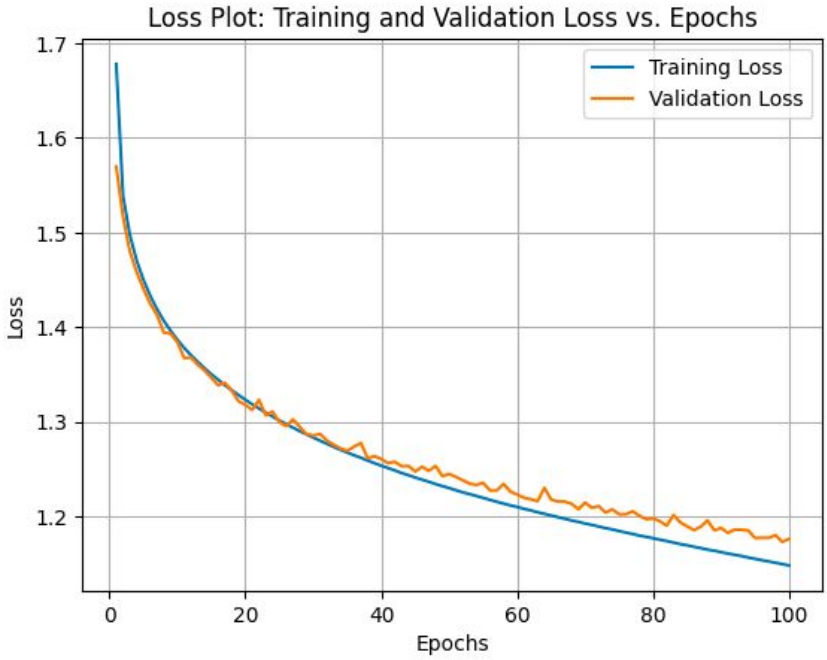

Figure 2: Model 2: Training Loss and Validation Loss

WHICH IS BETTER?

Model 3 is clearly better than model 4 because BERT is a better contextualized embedding than FastText. It produces a richer representation of the text as it uses a transformer architecture(where attention gives the required weight to the relevant words related to the current). Also BERT is better than GRU in doing Sequence tasks because it is a transformer hence model 3 is better than model 4.

TASK 2: EFR Task [Emotion Flip Reasoning]

MODEL 3 (Intuition and Architectur)

For each utterance, we computed the BERT embedding. Now the last utterance (say uL) in the dialogue is the one whose triggers need to be found. So we also find the previous utterance (uP) by that same speaker(who said the last utterance in the dialogue). In case the emotions are different then there is a flip. Now these two utterances are important so we concatenate these two. (uL + uP) Now this concatenated utterance is further concatenated with each utterance in the dialogue.

(ui = ui + (uL + uP))

Now each utterance has a context of the two important utterances and we need to find whether the current utterance is a trigger for the change in emotion. This is a binary classification task and again BERT for Sequence Classification is used for this.

MODEL 3 (Results)

M3 - Average Test Loss: 0.0039, Test accuracy: 0.8532, Test F1 score: 0.7856

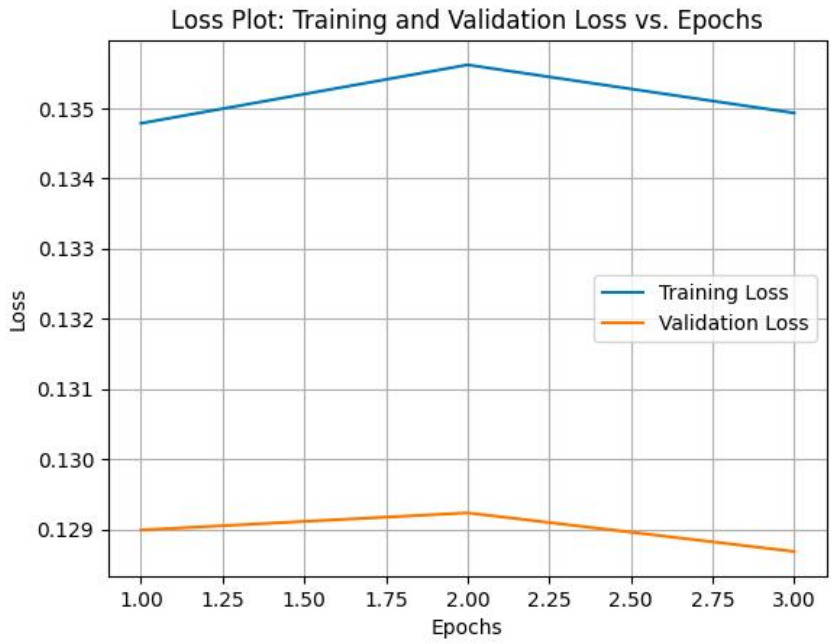

Figure 3: Model 3: Training Loss and Validation Loss

MODEL 4 (Intuition and Architecture)

This model is similar to model 3 and model 2 where we use fasttext representation for each utterance. We concatenate the embedding of the last utterance and previous utterance of the same speaker with every utterance. Now all these modified utterance embeddings are fed into a GRU and a binary classification task is done.

MODEL 4 (Results)

M4 - Average Test Loss: 0.0130, Test accuracy: 0.8532, Test F1 score: 0.7856

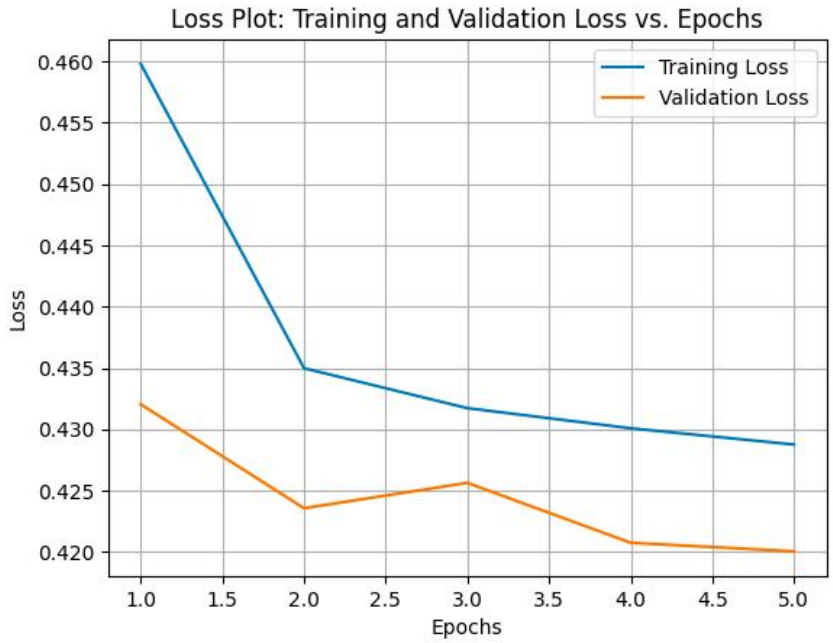

Figure 4: Model 4: Training Loss and Validation Loss

WHICH IS BETTER?

Model 3 is clearly better than model 4 because BERT is a better contextualized embedding than FastText. It produces a richer representation of the text as it uses a transformer architecture(where attention gives the required weight to the relevant words related to the current). Also BERT is better than GRU in doing Sequence tasks because it is a transformer hence model 3 is better than model 4.

Credit Statement

- **Arjit Singh Arora** - Task 1
- **Akshat Gupta** - Task 2
- **Kumar Aryan Singh** - Task 1
- **Swati Sharma** - Task 2