CSE 556: Natural Language Processing
Assignment 4

Date: April 11, 2024
Due Date: 11:59:59 pm, April 18, 2024 Max Marks: 100

General Instructions:

- Every assignment has to be attempted by four people. At least one subtask has to be done by one team member. All members need to have a working understanding of the entire code and assignment.
- Institute policies will apply in cases of plagiarism
- Create separate .ipynb or .py files for each part. The file name should follow the format: “A4_<Part number>.ipynb/.py”
- Create a single .ipynb file to generate the final outputs that are required for submittables. It should be named as “A4_<Group No>_infer.ipynb”. Clearly indicate which cell corresponds to the output of which task/subtask. Outputs will be checked from this inference file only by TAs.
- Carefully read the deliverables for all tasks. Along with the code files, submit all the other files mentioned for each task, strictly following the naming convention instructed.
- Only one person has to submit the zip file containing all the mentioned files and the report PDF. It will be named “A4_<Group No>.zip”. The person with the alphabetically smallest name should submit it.
- You are required to submit your trained models. You must also retain all your checkpoints and load and run them during the demo.
- Your report must include the details of each group member’s contribution.

This assignment is along the lines of the paper “Discovering Emotion and Reasoning its Flip in Multi-Party Conversations using Masked Memory Network and Transformer”

Datasets

The paper proposed a new dataset known as the MELD-FR for the novel task. We have modified the datasets that can be used to split for the train and val.

For testing purposes, your respective TA’s will be given separate chunks of modified data and the final evaluation will be the part of demos.

TASK 1 - ERC (Emotion Recognition in conversation)

What is ERC ?

Emotion Recognition in Conversation (ERC) is a specialized field that focuses on automatically identifying and interpreting the emotional states expressed by individuals during conversations. Unlike traditional approaches that analyze emotions in isolated text, ERC aims to understand the nuanced emotional dynamics in conversational exchanges involving multiple speakers. It enables a deeper understanding of emotional processes and interactions, benefiting applications such as conversational agents and affective computing systems.

What are you expected to do?

1. Train 2 models M1 & M2 with independent architectures which should be different from the architecture mentioned in the paper given above. (Note : If you are re-implementing the paper's model architecture as one of the models, make sure that the papers results will be considered as the baselines and the other model's scores should be at par with the one's given in the paper)

2. Both these models should focus on emotion identification and also detecting the flips in emotion within a set of conversations. Reasoning of the shift is not important for this part of the assignment.

3. Submit both model checkpoints M1 and M2, and report the better model with proper reasoning. The same will be tested on the testing data which will be provided at the time of demos.

Deliverables

1. Two model checkpoints M1 and M2 in proper format.[5*2=10]
2. Well labeled model architectures used for both M1 and M2.[5*2=10] (If implementing the paper's model, ignore this deliverable for only one model but note the baselines.)
3. If you are implementing the paper's architecture, then the other model should give results around at most 5% less than the paper's results. It can defeat the paper's results. [5 for comparable result on new model,5 for model architecture of of new model]
4. Properly mention which of the two architectures was better and why.[5]
5. Proper report explaining the intuition behind the models, splits and everything relevant.[5]
6. Add train loss and val loss vs epochs plots for each model in the report. [2.5*2*2=10]
7. 10 Marks for relevant viva questions.[10]

You can choose either point 2 or 3.

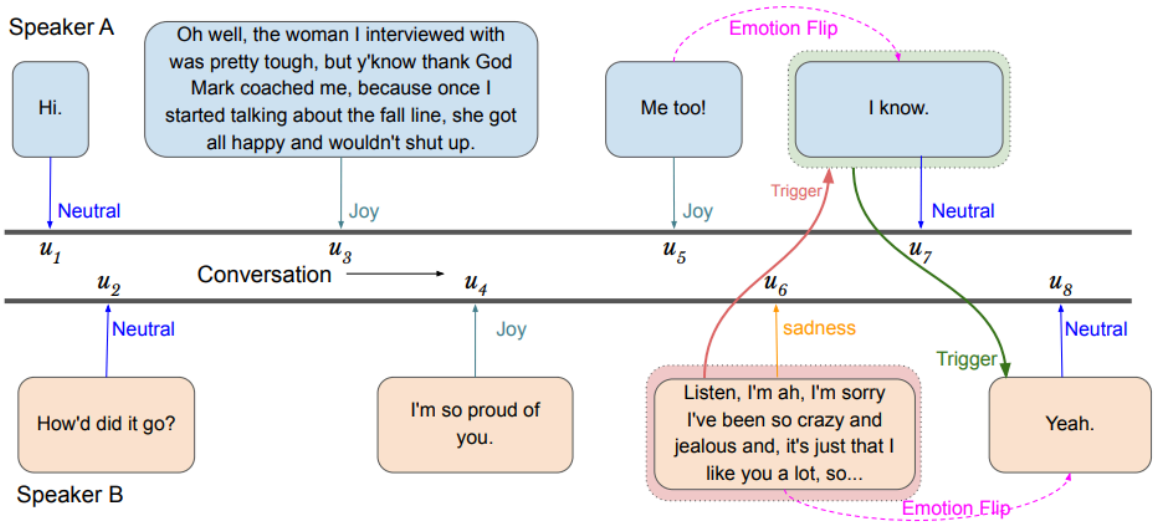
TASK 2 - EFR (Emotion Flip Reasoning)

What is EFR ?

In pursuit of this objective, the paper introduces a new challenge named EmotionFlip Reasoning (EFR) in conversational analysis. This task aims to identify all utterances within a dialogue that cause a speaker's emotional state to change. It is important to note that while some emotional shifts may be prompted by other speakers, there are instances where an individual's own utterance can serve as the catalyst for this change.

Formally, the EFR task can be defined as follows: Let $U = [u_1, u_2, ..., u_n]$ be a sequence of n utterances in a dialogue conversation D uttered by m distinct speakers $S = \{s_1, s_2, ..., s_m\}$. As the conversation progresses within a dialogue, these speakers utter their views or feelings in response to the previous utterances. For each utterance $u_i^{s_j}$ in the dialogue, we have an associated emotion $E_i^{s_j} \in \{anger, fear, disgust, sad, happy, surprise, neutral\}$, i.e., $E_i^{s_j} = f_E(u_i^{s_j})$. Consequently, if the emotion expressed in utterance $u_i^{s_j}$ flips w.r.t the speaker s_j's last utterance $u_l^{s_j}$, there might be a set of associated trigger-utterances u_k, $1 \leq k \leq i$, responsible for the emotion-flip of s_j, i.e., $[..., u_k, ...] = f_T(u_i^{s_j})$. In case of no emotion-flip, we associate a '*non-trigger*' label to the current utterance.

A snapshot from the paper itself



In the above mentioned example note the emotional flip in a set of conversations between multiple individuals. The task is to note the emotional shift. The purple arrow in the example shows the switch in the emotion from Joy to Neutral and sadness to neutral in the second case.

What are you expected to do?

4. Train 2 models M3 & M4 with independent architectures which should be different from the architecture mentioned in the paper given above. (Note : If you are re-implementing the paper's model architecture, make sure that the papers results will be considered as the baselines and your scores should be at par with the one's given in the paper)

5. Submit both model checkpoints M3 and M4, and report the better model with proper reasoning. The same will be tested on the testing data which will be provided at the time of demos.

Deliverables

1. Two model checkpoints M3 and M4 in proper format.[5*2=10]
2. Well labeled model architectures used for both M3 and M4.[5*2=10] (If implementing the paper's model, ignore this deliverable for only one model . For the other model you will have to but note the baselines.)
3. If you are implementing the paper's architecture, then the other model should give results around at most 5% less than the paper's results. It can defeat the paper's results. [5 for comparable result on new model,5 for model architecture of of new model]

4. Properly mention which of the two architectures was better and why.[5]

5. Proper report explaining the intuition behind the models, splits and everything relevant.[5]
6. Add train loss and val loss vs epochs plots for each model in the report. [2.5*2*2=10]
7. 10 Marks for relevant viva questions.[10]

You can choose either point 2 or 3.