Emotion Cause extraction using graphs to capture the flow of the conversation

Task 1: Dataset Preparation

In the RECCON dataset, each conversation is labeled as tr_xxxx. Each conversation has labeled utterances and their causes. Your task is to first take each utterance and break it into clauses. For example: “Why? What happened?” Gives 2 clauses: “Why?” And “What happened?” Similarly, not just “?”, consider all delimiters, conjunctions, etc, to break these utterances into meaningful clauses.

Note: Any automated method for data collection carries more weightage than manually collecting them.

Task 2: Annotation

Use the dataset to label the clauses as

1. Emotion clause: the clause that shows the emotion.
2. Cause clause: the clause that shows the cause.

Note: A clause can be both an emotion clause and a cause clause. Again automated method here is appreciated.

Task 3: Model

Given a conversation, first derive different clauses present in the utterance. Then, text embedding for each clause is obtained. Each clause becomes a node. Based on your annotation, each node in the graph is an emotion, a cause, or neither**. Find a way to connect these nodes with a property as an edge** (e.g., subject-verb, verb-object). Feed this graph to DyGCN and learn parameters to predict the emotion-cause pair.