NLP Project: Textual Emotion-Cause Pair Extraction in Conversations
Group - 42

Ritwik Harit	Vasan Vohra	Sachin Sharma	Tanmay Singh
3rd Year B.Tech CSAI	*3rd Year B.Tech CSAI*	*3rd Year B.Tech CSAI*	*3rd Year B.Tech CSAI*
IIIT Delhi	*IIIT Delhi*	*IIIT Delhi*	*IIIT Delhi*
Delhi, India	Delhi, India	Delhi, India	Delhi, India
Roll no. 2021557	Roll no. 2021572	Roll no. 2021559	Roll no. 2021569

INTRODUCTION

In the domain of textual emotion analysis, much of the research has been centered on recognizing emotions conveyed through text. However, there has been a significant gap in understanding the hidden causes of these emotions, specifically within the context of conversational exchanges. The existing studies and scholarly works have focused mainly on static textual sources such as social media posts or news articles. Yet, the intricate dynamics of real-time conversations have not been addressed and remain relatively unexplored. Like all natural forms of human interaction, conversations encompass textual elements and non-verbal cues such as the tone of the voice and facial expressions. The multimodal features within the conversational contexts enhance our understanding of the hidden emotions behind conversations. Although previous studies have accounted for multimodal emotion analysis within conversations, the origin of emotions needs to be more explicitly included in studies. Thus, in our study, Textual Emotion-Cause Pair Extraction in Conversations, we aim to explain the causal factors underlying expressed emotions in textual dialogue. The primary objective of our study is to analyze conversations to identify instances that acted as triggers for the corresponding articulated emotions. For our study, we used a dataset derived from conversations in the renowned television series "Friends." The corpus contains substantial conversational exchanges carefully annotated to denote instances of expressed emotions and their associated causes. To showcase how emotion-cause pairs are manifested within a conversational discourse, we use an illustrative example from the dataset to demonstrate our approach. We developed a sophisticated system to ensure the extraction of emotion-cause pairs from textual conversations, incorporating advanced techniques to parse and analyze the intricacies of human expression. The experimental results of our study demonstrate the effect of efficient extraction and representation of utterances for discovering emotions and causes in conversations.

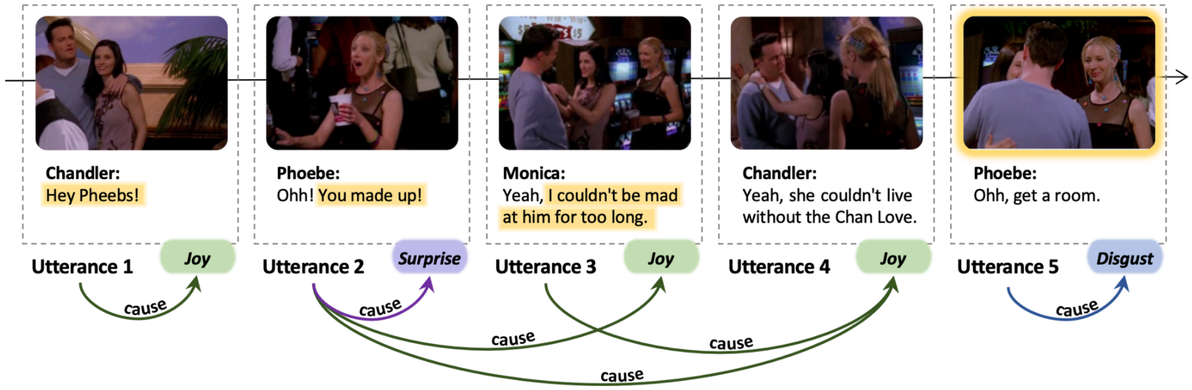


Fig: An illustrative example of an annotated conversation from the Emotion-Cause-in-Friends (ECF) dataset. Each utterance has an associated emotion attached to it, and every emotion is triggered by a specific 'span' within the cause/trigger sentence.

RELATED WORK

Emotion cause analysis is an emerging and rapidly expanding field of study that has garnered significant attention across various academic domains due to its practical applications, such as improving product quality [1] and analyzing public opinions [2], [3].

The groundwork for defining emotion cause analysis was established by Lee et al. [4]. Since then, several methods have been proposed to address this task. Gao et al. [5] developed an approach to detect emotion causes in microblogs using the Emotion-Cause-OCC (ECOCC) model. Gui et al. [6] introduced a method based on convolutional kernels and multi-kernel SVM, trained on publicly available datasets from SINA city news for emotion cause analysis. Li et al. [7] devised a decision-tree method that relies on specific words as clues. Han et al. [8] utilized a structural SVM model and incorporated features from various linguistic elements like words, distances, context, and language patterns.

Despite these advancements, current methods heavily depend on manually creating features, which becomes increasingly time-consuming as the volume of data grows. Additionally, the diverse structures of clauses pose challenges in manually extracting features, highlighting the necessity for more efficient and automated approaches to feature extraction.
Building upon Talmy's framework (2000), the cause of an emotion is conceptualized as an event itself, termed here as a "cause event." When referring to cause events, it is essential to clarify that we are not referring to the precise trigger of the emotion or its direct precursor. Instead, we focus on the immediate cause of the emotion, which can encompass either the actual triggering event or the individual's perception of said event. Drawing from the TimeML annotation scheme (Saurí et al., 2004), *events* are defined as occurrences or situations. In this study, cause events denote

the explicitly stated arguments or events closely associated with the manifestation of corresponding emotions.

In Lee et al.'s (2010) corpus, cause events are classified into two primary categories: verbal and nominal events. Verbal events encompass actions involving verbs, including propositions and nominalizations, while nominal events consist solely of nouns or nominals.
In recent years, there has been considerable focus on exploring the integration of sentiment classification with emotion classification [9,10,11] and sentiment classification with sentiment topic detection [12, 13]. These studies have shown promising results, capitalizing on the inherent connections between these subtasks. Motivated by these findings, some researchers have endeavoured to classify the emotional polarity of cause representations, often incorporating emotional tendencies, to enhance the Emotion-Cause Identification (ECI) task.

METHODOLOGY

For our study, we utilized the Emotion-Cause-in-Friends (ECF) dataset, which consists of a substantial amount of conversations between the show's characters, each having multiple sentences. We approach our study by breaking the problem statement into two specific tasks. Task 1 is based on identifying the underlying emotions within a conversation, i.e., Emotion Recognition in Conversation. Task 1 is further subdivided into two separate models accounting for the absence and presence of previous contexts while identifying emotions. For model 1, we accounted for the speakers and utilized the context of the conversation. We created context embeddings by concatenating the context of an utterance in its conversation and obtained contextual embeddings from BERT. We added the speaker names before the sentences to capture the speaker context and passed the context embeddings to the BERT encoder, BertForSequenceClassification model.

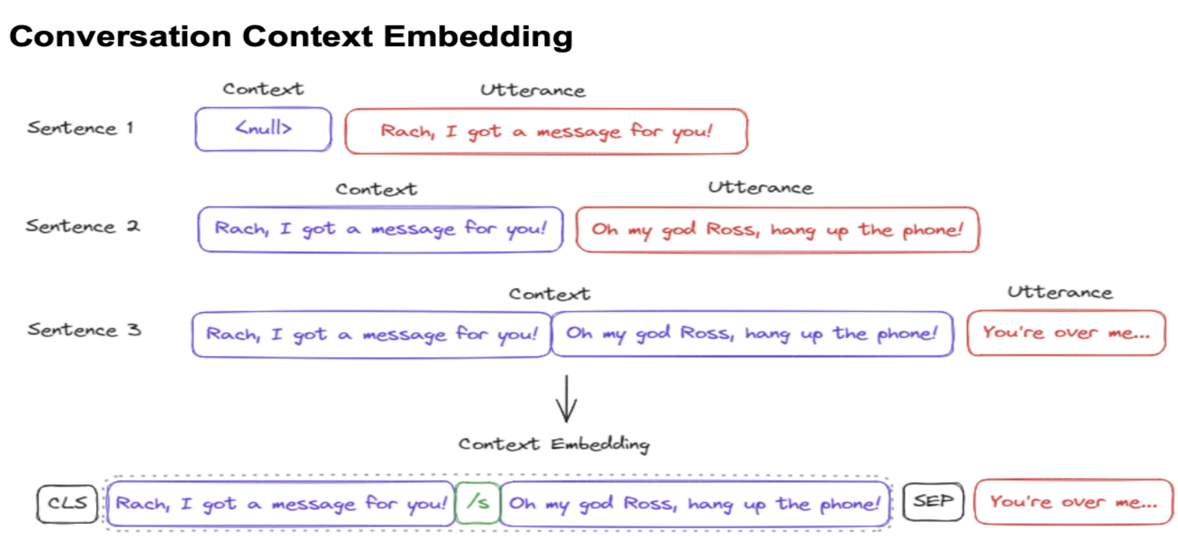


Fig: An illustration of Model 1 for Task 1 (BERT with Conversational Context Embeddings)

For model 2, rather than concatenating previous utterances in a conversation to provide context, we let the Bert model itself learn the context by first predicting emotions for all utterances in conversation, and then backpropagating the combined loss, enabling Bert to optimize emotions for all utterances in the conversation at once. We added the speaker names before the sentences to capture the speaker context and passed the context embeddings, generated using BERT Tokenizer, to the BERT encoder, BertForSequenceClassification model and update weights with a combined loss of all utterances of a conversation at once.

While task 2 is based on finding the emotion-cause pairs within a conversation, i.e., identifying the triggers(causes) that led to a specific emotion within the context of a conversation. For Task 2, we developed two different models to perform the above-mentioned task. Model 1 uses a fine tuned pre-trained BERT for question answering to identify the span, while Model 2 uses a fine tuned pre trained BERT model to extract the target-cause/trigger pairs that led to the manifestation of a specific emotion. Once the cause pairs are identified, we look for the specific portion within the sentence, the 'span,' which acted as the primary trigger for expressing emotion, using another BERT model.

DATASET DESCRIPTION

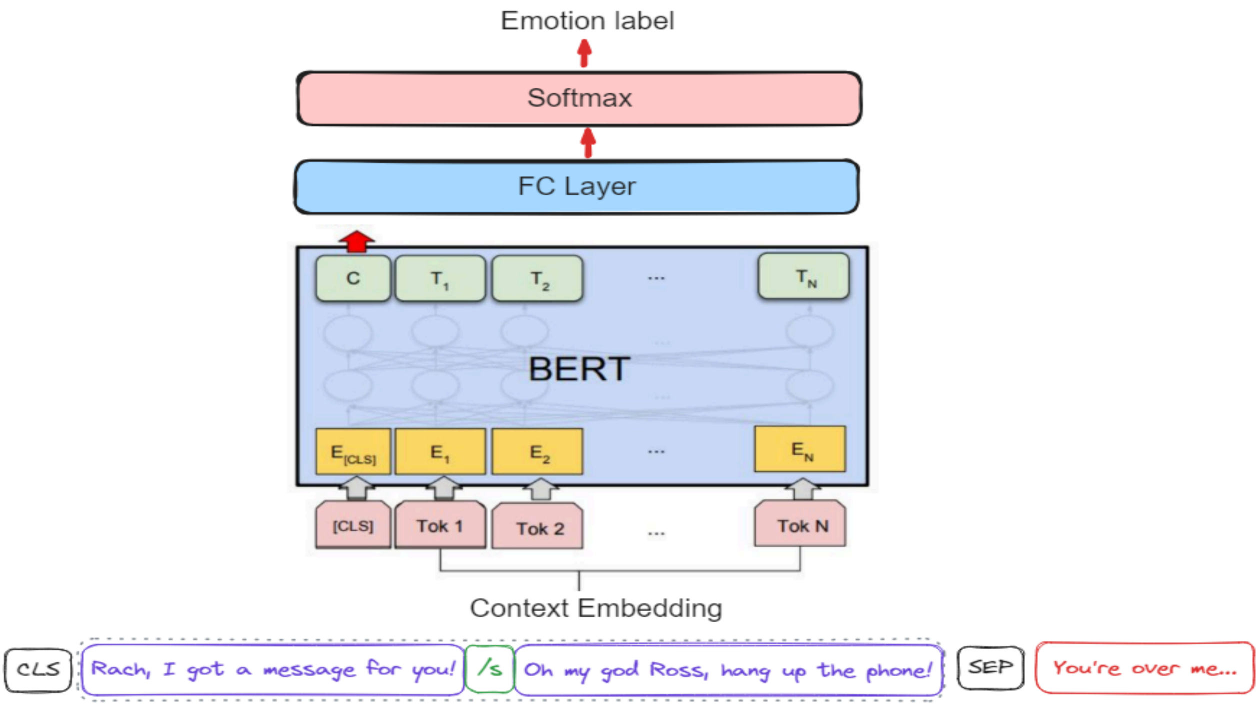
In our study, we used the Emotion-Cause-in-Friends (ECF) dataset, which leverages the famous sitcom Friends as its primary source material. Within the ECF dataset, researchers have meticulously curated 1,344 conversations and meticulously annotated 13,509 utterances. This comprehensive dataset encompasses a staggering 9,272 emotion-cause pairs meticulously annotated across three distinct modalities. Such an expansive collection provides researchers with a rich resource for exploring and understanding the intricate interplay between emotions and their underlying causes within everyday social interactions, as portrayed in the Friends series.

EXPERIMENTAL SETUP

ERC Model 1: BERT with Conversation Context Embeddings

Each input embedding is of size 512. Bert Model consists of:

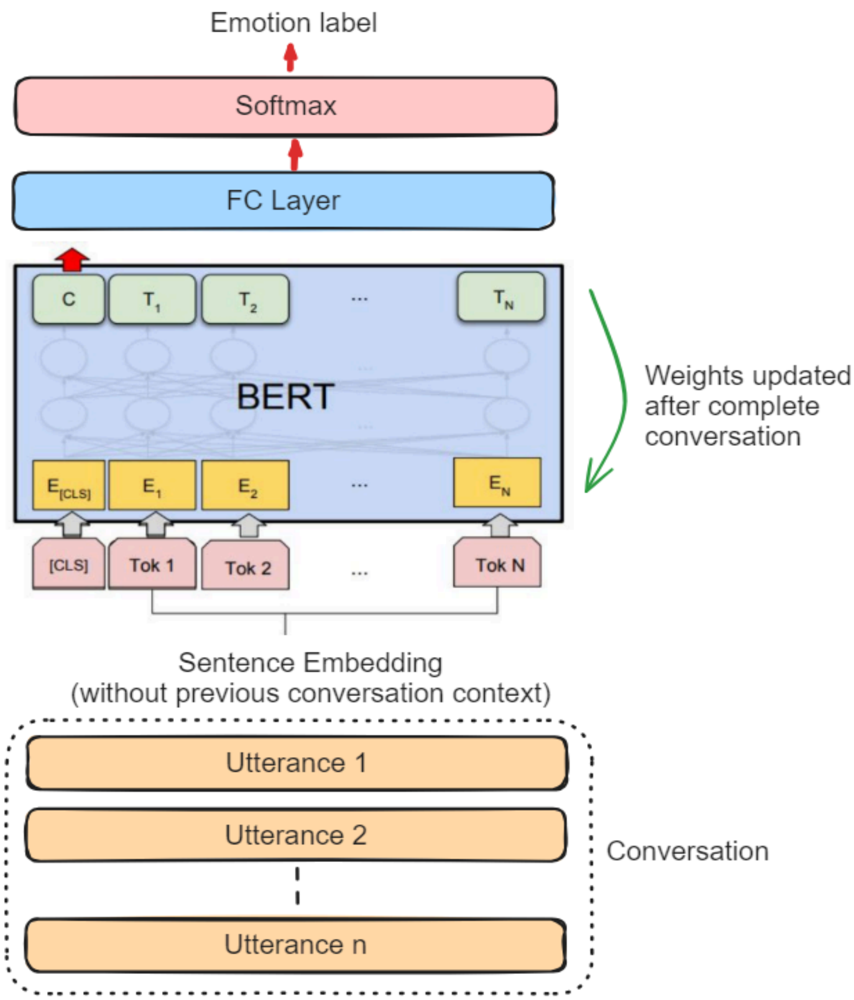
- Bert Embedding: Token + Positional
- Bert Encoder: Self Attention + Add & Norm + FC
- Pooling
- Classifier: Dense + Softmax



ERC Model 2: BERT without Conversation Context Embeddings

Each input embedding is of size 512. Bert Model consists of:

- Bert Embedding: Token + Positional
- Bert Encoder: Self Attention + Add & Norm + FC
- Pooling
- Classifier: Dense + Softmax

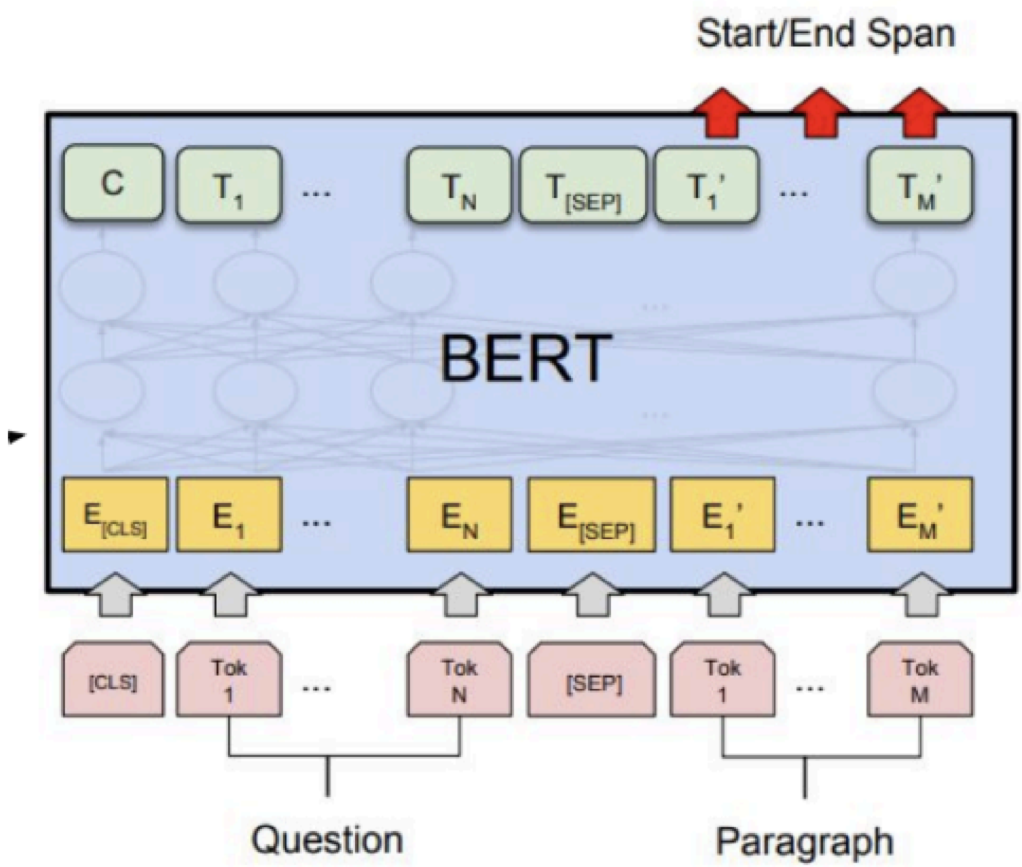


TASK 2

Model 1

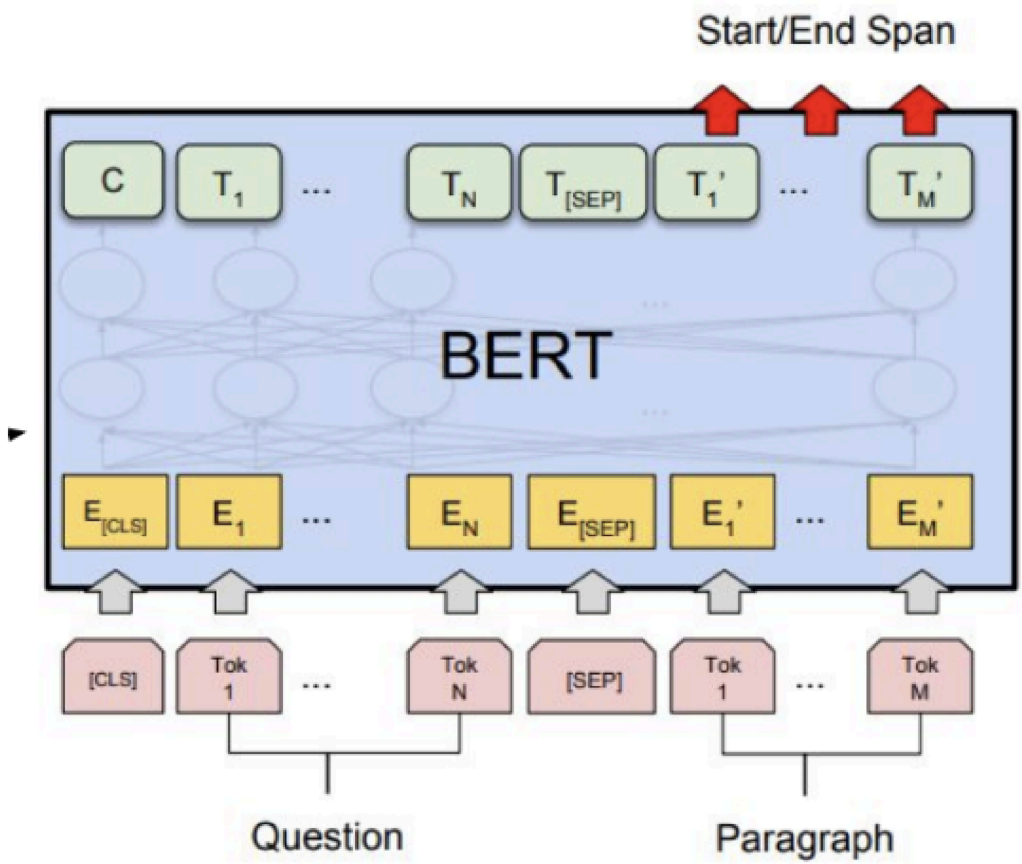
- We fine-tuned the pre-trained BERT model as a Question-Answering Model
- The target (emotion manifestation) has a causal utterance in conversational contexts

- Each causal utterance has a span that acts as a cause for a specific emotion
- Out of all context utterances, we try to directly identify what span within an utterance can be a cause for an emotion



Model 2

- We fine-tuned the pre-trained BERT model as a Question-Answering Model
- The target (emotion manifestation) has a causal utterance in conversational contexts
- Each causal utterance has a span that acts as a cause for a specific emotion
- Out of all context utterances, we try to identify the utterances which can act as a cause for a specific emotion.
- After identifying the causal utterances, we extract the span within the utterance that acts as the primary cause for a given emotion.



RESULTS

TASK 1

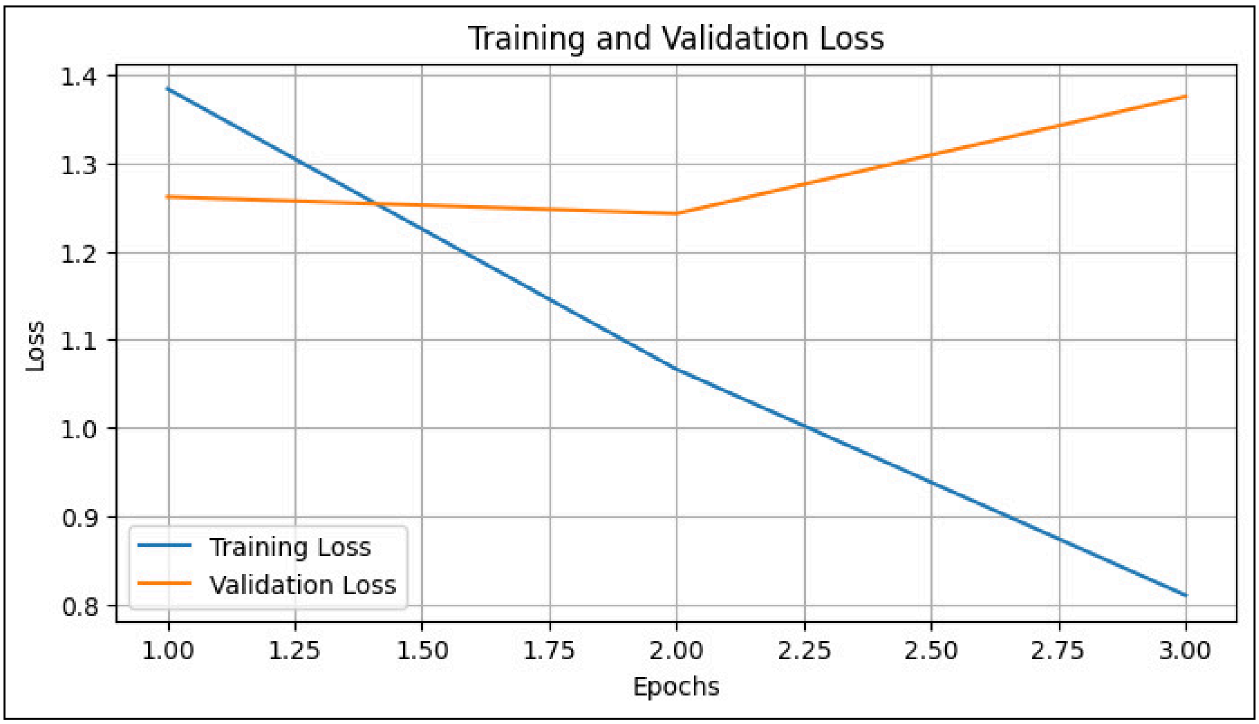
ERC Model 1 Results

	Precision	Recall	F1-Score	Support
Anger	0.45	0.25	0.32	294
Disgust	0.37	0.07	0.12	96
Fear	0.25	0.14	0.18	90
Joy	0.53	0.49	0.51	485
Neutral	0.63	0.71	0.67	1132
Sadness	0.39	0.42	0.40	198
Surprise	0.51	0.70	0.59	359

Accuracy			0.55	2654
Macro Avg	0.45	0.40	0.40	2654
Weighted Avg	0.54	0.55	0.54	2654

Accuracy: 0.5535041446872645
Precision: 0.5372222458617016
Recall: 0.5535041446872645
F1 Score: 0.5353365847232632

Loss Plot :

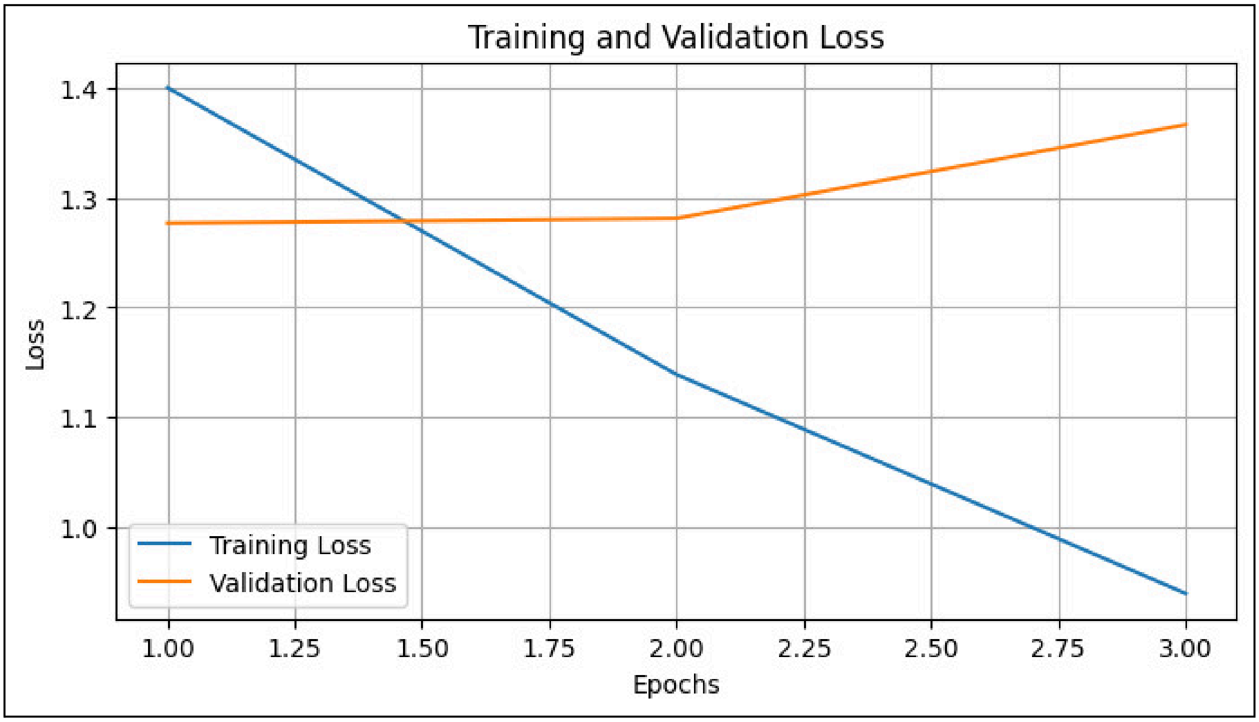


ERC Model 2 Results

	Precision	Recall	F1-Score	Support
Anger	0.39	0.34	0.36	294
Disgust	0.33	0.07	0.12	96
Fear	0.07	0.02	0.03	90
Joy	0.52	0.49	0.50	485
Neutral	0.63	0.70	0.66	1132
Sadness	0.28	0.42	0.34	198
Surprise	0.58	0.55	0.56	359

Accuracy			0.53	2654
Macro Avg	0.40	0.37	0.37	2654
Weighted Avg	0.52	0.53	0.52	2654

Accuracy: 0.53353428786737
Precision: 0.5200249251666161
Recall: 0.53353428786737
F1 Score: 0.5209382264091214



TASK 2

Span Identification for Model 1 results

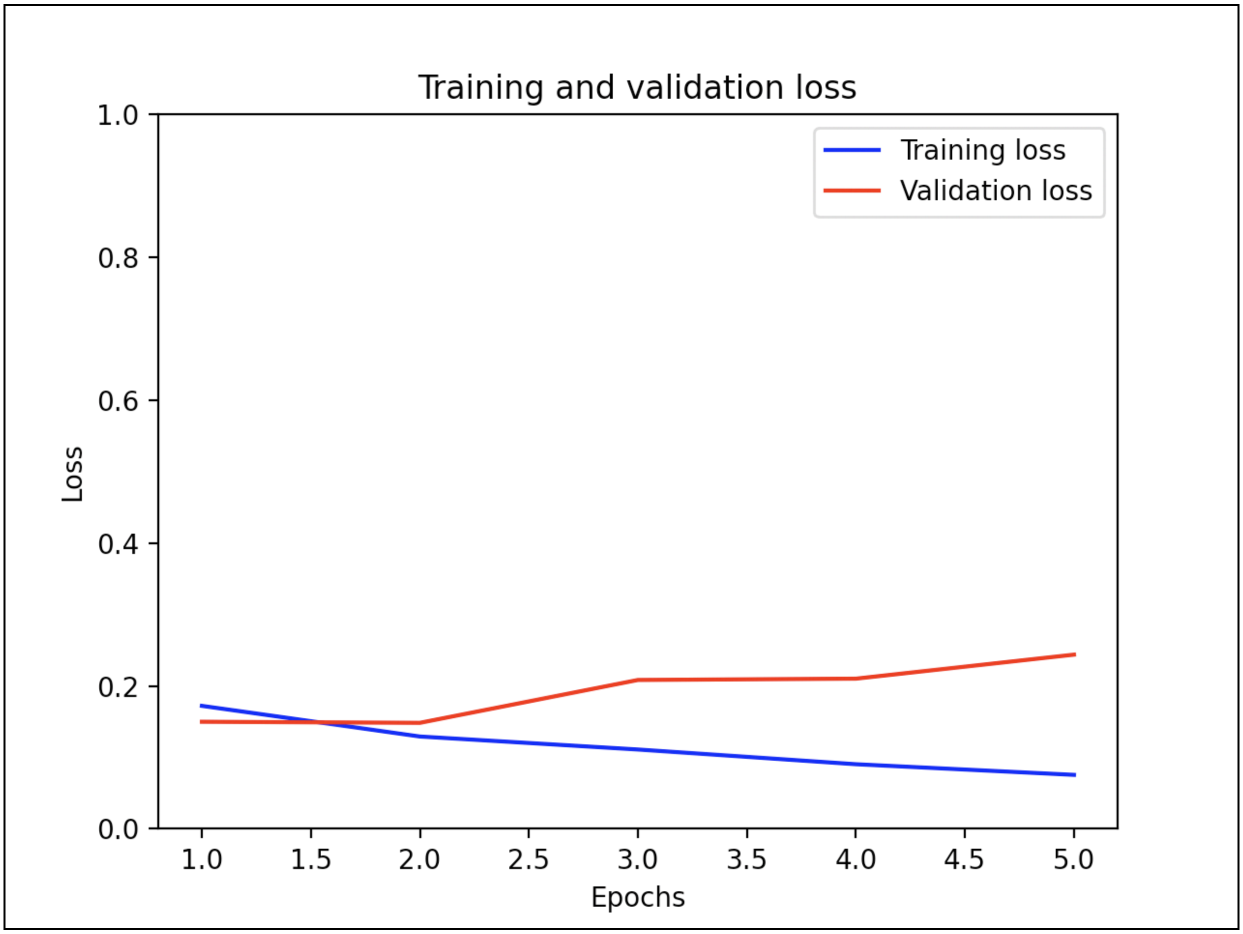
Accuracy: 0.9427426281133696
Precision: 0.9614825218976901
Recall: 0.9427426281133696
F1 Score: 0.9336956203820744

	Bleu 1	**Bleu 2**	**Bleu 3**	**Bleu 4**
Bleu Score	0.8874815771608535	0.8823986262743536	0.8712855783701619	0.71398371524857

Precisions	[0.9265052966683133]	[0.9265052966683133, 0.9159227790922779]	[0.9265052966683133, 0.9159227790922779, 0.8868302326760278]	[0.9265052966683133, 0.9159227790922779, 0.8868302326760278, 0.4101665205959685]
Brevity Penalty	0.9578807378136014	0.9578807378136014	0.9578807378136014	0.9578807378136014
Length Ratio	0.9587433563121612	0.9587433563121612	0.9587433563121612	0.9587433563121612
Translation Length	109314	109314	109314	109314
Reference Length	114018	114018	114018	114018

Meteor Score : 0.9299682500960135

Loss Plot :



Utterance Identification for Model 2 Results

	Precision	Recall	F1-Score	Support
0	0.97	0.98	0.97	33866
1	0.57	0.38	0.45	1892
Accuracy			0.95	35784
Macro Avg	0.77	0.68	0.71	35784
Weighted Avg	0.94	0.95	0.95	35784

Accuracy: 0.9517310811566643
Precision: 0.9446621989820667
Recall: 0.9517310811566643

F1 Score: 0.9471150524657891

Span Identification for Model 2 results

	Bleu 1	**Bleu 2**	**Bleu 3**	**Bleu 4**
Bleu Score	0.7939387080020839	0.7871580816020524	0.7806733323525046	0.7736548445073504
Precisions	0.8949547615396412	[0.8949547615396412, 0.8797333333333334]	[0.8949547615396412, 0.8797333333333334, 0.8655620805369127]	[0.8949547615396412, 0.8797333333333334, 0.8655620805369127, 0.8487798710330003]
Brevity Penalty	0.8871271958330345	0.8871271958330345	0.8871271958330345	0.8871271958330345
Length Ratio	0.8930430019172829	0.8930430019172829	0.8930430019172829	0.8930430019172829
Translation Length	13042	13042	13042	13042
Reference Length	14604	14604	14604	14604

Meteor Score : 0.8392483788571296

DISCUSSION

From the results of our ERC model, we can observe that Fear & Disgust show the lowest F1 scores. The primary reason for lower f1 scores corresponding to these emotions is that the number of instances available in the corpus is lower in juxtaposition to other emotions.

Moreover, our model for utterance identification in Task 2 performs very well, as shown by the high weighted average F1-scores (approximately 0.95). However, one of the reasons behind such a high score is the presence of a very large number of 0s as compared to the number of 1s. Furthermore, the utterance pair identification task is very similar to sentence similarity matching, for which the BERT model performs remarkably well. Thus, fine-tuning the model for our use case has given rise to a powerful and robust model.

CONCLUSION

The primary objective of our study was to identify the emotion-cause pairs within the context of conversations, and our experiments yielded positive results, wherein we successfully identified the cause/trigger sentences that led to the manifestation of a specific emotion within a conversational discourse. After extracting the cause/trigger sentences, we successfully identified the 'span' within the sentence that acted as the primary source of trigger for expressing a corresponding emotion.

FUTURE WORK

For future work, we would like to fine-tune our model to ensure better performance and effectively model the impact of speaker relevance on emotion recognition and emotion-cause extraction in conversations.

REFERENCES

[1] Bell S., Bala K., 2015, Learning visual similarity for product design with convolutional neural networks

[2] Liu B., 2012, Sentiment Analysis and Opinion Mining

[3] Song K., Wang D., Feng S., Zhang Y., Qu W., Yu G., 2014, CTROF: A Collaborative tweet ranking framework for online personalized recommendation

[4] Lee S.Y.M., Chen Y., Huang C.-R., 2010, A text-driven rule-based system for emotion cause detection

[5] Gao K., Xu H., Wang J., 2015, A rule-based approach to emotion cause detection for chinese micro-blogs

[6] Gui L., Xu R., Lu Q., Wu D., Zhou Y., 2016, Emotion cause extraction, a challenging task with corpus construction

[7] X. Li, S. Feng, D. Wang, Y. Zhang, Decision tree method for the ntcir-13 eca task.

[8] H. Ren, Y. Ren, J. Wan, The gdufs system in ntcir-13 eca task.
[9] Gao W, Li S, Lee S, at al (2013) Joint learning on sentiment and emotion classification. In: Proceedings of the 22nd ACM international conference on information & knowledge management, pp 1505–1508
[10] Wang R, Li S, Zhou D, at al (2015) Joint sentiment and emotion classification with integer linear programming. In: Liu A, Ishikawa Y, Qian T, Nutanong S, Cheema M (eds) Database systems for advanced applications, lecture notes in computer science, Vol 9052
[11] You Q (2016) Sentiment and emotion analysis for social multimedia: methodologies and Applications[C]// the 2016 ACM ACM
[12] Pavitra R, Kalaivaani PCD (2015) Weakly supervised sentiment analysis using joint sentiment topic detection with bigrams[C]// 2015 2nd International Conference on Electronics and Communication Systems (ICECS) IEEE
[13] Huang F, Zhang S, Zhang J et al Multimodal learning for topic sentiment analysis in microblogging[J].