Textual Emotion-Cause Pair Extraction in Conversations

Abstract— **Dialogue is a vital aspect of human interactions that contains a significant number of emotions, making it intriguing to find their origins within conversations. In its organic state, dialogue encompasses various modes of expression, rendering it a complex and multifaceted form of communication. Importantly, conversations occur in a natural, multimodal manner, integrating various modes of expression for a comprehensive communication experience.**

1. Introduction

Emotion plays a vital role in human conversations. Our task is to predict the cause of a particular emotion in the conversations. Conversation in a natural form is multimodal. Multimodal is especially for discovering both emotions and their cause. That is why two types of models are combined, which would predict the emotion of the utterance and the cause of the utterance.

2. Related Work

There has been a lot of background work in recent years in the form of Emotion Cause Extraction(ECE) as well as Emotion Cause Pair Extraction(ECPE). Lee et al.[1] defined and developed a rule based emotion extraction. Another paper from Li et al.[2] uses co-attention along with convolutional layers to predict cause present in an utterance as a binary classification problem.

The derivative of this task is to find Emotion-Cause pairs without knowing the emotions beforehand. There are two approaches to this problem. One involves using emotion to extract Emotion-Cause pairs. Mathur et al.[3], using textual, visual and audio datasets, predicted the emotion of a sentence and then used fully-connected layers, BiLSTMs and CRF layers along with emotion predictions to achieve 3 models.

Second approach is to not predict the emotions of the sentences before-hand but to only use the sentences to extract Emotion-Cause Pairs. A brilliant model developed by Wang et al. [4] involves extracting emotion possessing and causal clauses using BiLSTMs and afterwards forming all possible Emotion-Cause pairs and predicting their validity. Another paper by Nguyen et al. [5] involves using a text to text model where the input is provided as a question to find a potential cause clause from a sentence given the emotion sentence as the context and the model is fine tuned to provide the right clause from the sentence.

3. Dataset

The dataset provided for Task 3 of SemEval-2024 involves Emotion-Cause Pair Extraction, where the goal is to identify pairs consisting of an emotion and its corresponding cause within the conversations. It consists of 1,374 total conversations. Each conversation contains a series of sentences, with the maximum length of sentences in a conversation being 33. The number of classes of emotions is 7. We divided the training dataset into training and validation. The bar graphs of count of each emotion in the training(Fig.1), validation(Fig.2) and testing(Fig.3) datasets are shown below. We can see that the number of sentences in neutral emotion is higher as compared to the rest of the emotions in each of the datasets. The train, validation and testing datasets also have almost an equal proportion of the number of sentences in each emotion. The amount of sentences in the class of “disgust” and “fear” is low compared to other emotions which would probably lead to poor performance in those classes.

Each emotion-cause pair consists of an emotion and its corresponding cause within the conversation and 9,794 such emotion-cause pairs are shown. The dataset may contain future references which means that the cause of an emotion would come after the emotion. Also, some conversations may not contain any emotion-cause pairs.

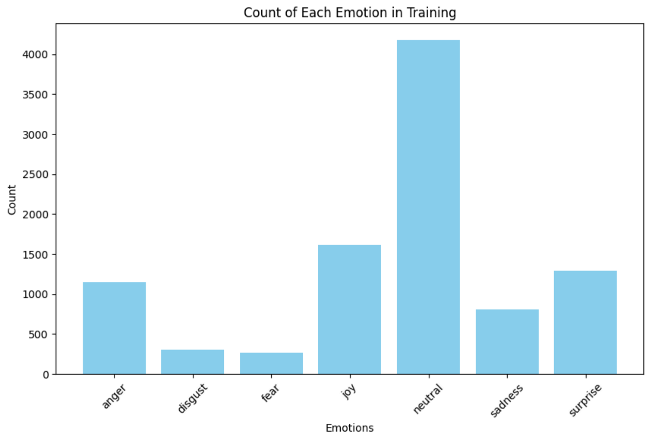

Figure 1: The count of the number of sentences in each emotion in the training dataset.

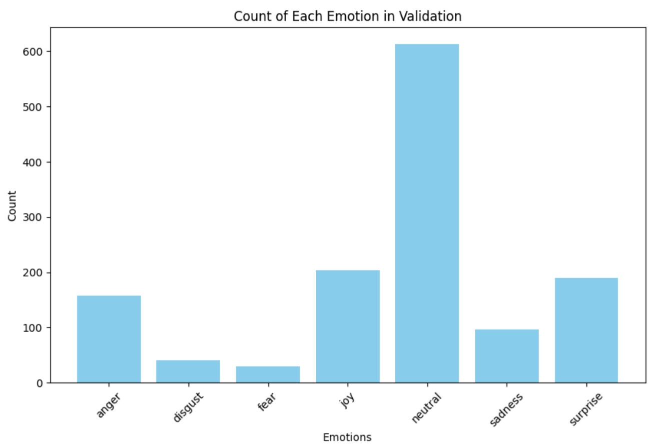

Figure 2: The count of the number of sentences in each emotion in the validation dataset.

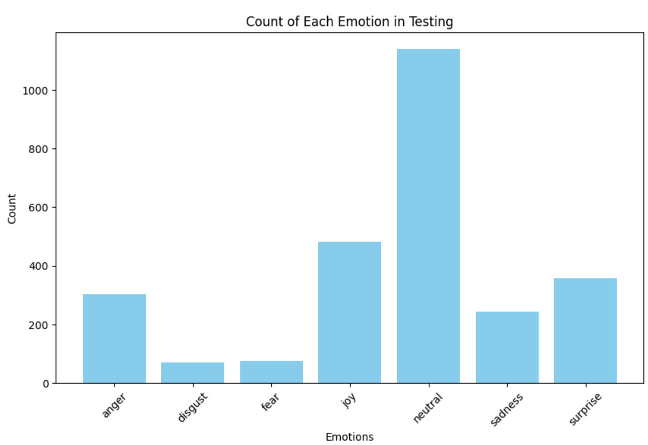

Figure 3: The count of the number of sentences in each emotion in the testing dataset.

4. METHODOLOGY

We define the ECPE task as: Given a conversation, predict all Emotion-Cause Pairs along with the emotion in the Emotion Sentence. Each pair consists of a sentence containing the emotion as well as a cause clause that is the reason for that emotion. We propose a setup involving 4 different models.

Before any training, we first preprocessed the conversations using SentenceBert[6]. The advantage of using SentenceBert is that we can get sentence-level representations from a pre-trained model is that it is fast and we may not require another layer to represent sentences into vectors as Wang et al.[4] did in their setup. We also padded the conversations with dummy sentences to make all conversations of equal length.

4.1 Model 1: Emotion Cause Extraction

The purpose of this model is to find all potential emotion possessing and causal sentences of each conversation. For this task, we propose two BiLSTMs. One for emotion sentences extraction and other for cause sentences extraction. Each BiLSTM is followed by a fully connected layer (3 linear layers, each followed by a leaky ReLU() except the last one). In this process, we use the SentenceBert representations as well we one-hot encode the speakers. We concatenate every sentence vector with its corresponding one-hot encoded speaker. The purpose of including speakers is to make dependency on the speaker as well. We send all these vectors to both the BiLSTMs. After getting certain representations of these input vectors, we send these representations to the fully connected layer in order to classify the representations as 0 or 1 respectively[4]. Our vision here is that, one BiLSTM with a classification head tries to predict whether a sentence contains an emotion or not and the other BiLSTM with a classification tries to predict whether a sentence contains a cause or not. This loss function for both the BiLSTMs is Cross Entropy Loss. In addition, we add both the losses and then propagate with the combined loss. This adds dependency of both BiLSTMs on each other. To further elucidate this, we believe that causes help identify emotions better and emotions help identify causes better and therefore the combination of losses. After training, the inferences with label 1 are our emotions containing representations and causal representations respectively. The optimizer used is Adam.

4.2 Model 2: Emotion Pair Classifier

This model utilizes the emotion and causal representations inferred from Model 1. The idea of this model is to recognise relationships between emotion representations and causal representations. Here, we want to identify all correct pairs of emotions and causal representations using both emotion and causal sets obtained previously. We therefore form all possible pairs of emotions and causes within each conversation and train a classifier to predict all valid pairs as a binary classification where 1 corresponds to a valid pair and 0 corresponds to an invalid pair. The classifier consists of a few Linear layers each followed by a Leaky ReLU layer except the last layer. The final layer predicts whether a pair is valid or not and the pairs with prediction 1 are passed on to the next model. The loss function used is Cross Entropy Loss and Optimizer as Adam.

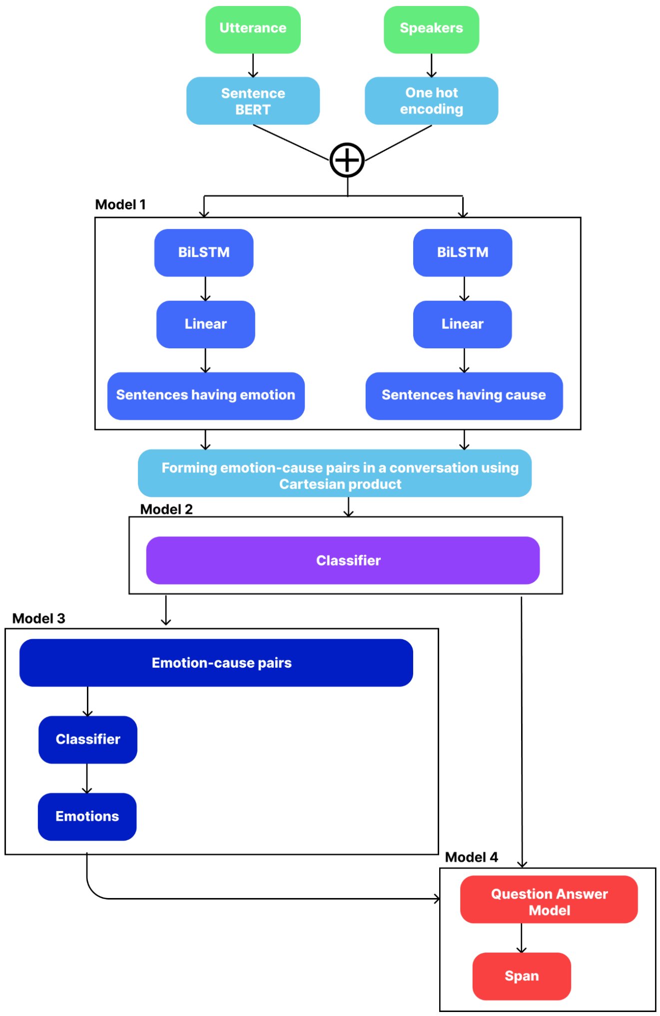

Figure 4: The combined model for ECPE.

4.3 Model 3: Emotion Classifier

After getting valid emotion-cause pairs, we finally train another classifier that utilizes the emotion-cause pairs to predict the emotion involved in those pairs. We take 6 emotions and 1 neutral emotion as our target and then train a classifier with 5 linear layers, each followed by a Leaky ReLU except the last layer. The last layer predicts the emotion and this model uses Cross Entropy Loss and Adam Optimizer. The pairs as well as their emotions are further passed on to the next model.

4.4 Model 4: Span Identification

For the final part of the model, a pre-trained question-answer model is used. At this moment, the whole sentence is being predicted which is the cause of the target sentence. Now to find the particular span (substring of the cause sentence), the question-answer model is used. The question is a sentence in the format of "Which part of the context is the cause for the utterance <target-sentence> with an <emotion-of-the-sentence> emotion?". That is how the emotional part is being incorporated into the question. Then the cause sentence is given as the context. The question-answer model will output the suitable span of the sentence.

5. Experimental Setup

SentenceBert takes in a sentence as a string and outputs a 384-dimensional vector. Model 1 is then trained for 10 epochs and the results and metrics are calculated. The learning rate for Model 1 for Adam Optimizer is 10^{-3}. Model 2 is trained for 15 epochs and learning rate for Adam Optimizer is 10^{-5}. Model 3 is trained for 10 epochs and learning rate for the Adam Optimizer is 10^{-5}.

6. Results and Analysis

Task of the model	F1_score
Model 1: Does not have emotion	0.61
Model 1: Has emotion	0.73
Model 1: Is not a cause	0.69
Model 1: Is a cause	0.63
Model 2: Is not a valid emotion-cause pair	0.93
Model 2: Is a valid emotion-cause pair	0.40
Model 3: disgust	0.00
Model 3: joy	0.52
Model 3: surprise	0.41
Model 3: anger	0.37
Model 3: fear	0.00
Model 3: sadness	0.27
Model 4: Final span for emotion cause-pair	0.0268

Table 1: F1 score for the output of different model

Our evaluation metrics consist of F1 score. F1 score incorporates both Precision and Recall. Model 1 provides a

decent score. The F1-score for the model 2, predicting if an emotion-cause pair is valid or not is low because a sentence can have multiple clauses, and each clause can denote a different cause which reduces its probability of being classified as a valid emotion-cause pair. However, from the paper by Wang et al.[4], all metrics are still near to their metrics. For emotion recognition, we were able to beat some of their emotions like joy, sadness, surprise.
On further analyzing the errors of emotions, we described how there are very few samples of fear and disgust which may not be very predictive by the models. This score further provides the evidence for the same.

7. Conclusion And Future Work

In this paper, we described an ECPE task where we try to predict all emotion cause pairs. This task comes from Task 3 of SemEval-2024. Furthermore, we opted not to use emotions or predict emotions in order to solve the task as ECPE involves predictions without the emotions as input. We use BiLSTMs to extract emotion and cause representations. We then train classifiers to classify emotion cause pairs formed by cartesian product of emotion set and cause set within each conversation. We further use those pairs to predict the emotions as well and then use a Question-Answering Model in order to extract the span from cause sentences.
The extracted emotion-cause pairs can be used in various applications, such as sentiment analysis, customer feedback analysis, and dialogue systems. Understanding the causes of emotions can help improve communication and empathy in human-computer interactions.
Furthermore, it is worth noting that ECPE is a challenging task as compared to ECE. Our model was built in order to work on sentences rather than clauses which gives worse performances because sentences contain multiple clauses. Moreover, our model performance is justified due to the fact that we train multiple models and these models have some dependencies on one other and therefore, errors of one model are propagated to the other. As a result, we call for future work in leveraging end-to-end architectures in order to reduce error propagation as well as opt for better designs and methods for span identification.

8. Reference

https://nustm.github.io/SemEval-2024_ECAC/

[1] Sophia Yat Mei Lee, Ying Chen, and Chu-Ren Huang. 2010. A Text-driven Rule-based System for Emotion Cause Detection. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 45–53, Los Angeles, CA. Association for Computational Linguistics.

[2] Xiangju Li, Kaisong Song, Shi Feng, Daling Wang, and Yifei Zhang. 2018. A Co-Attention Neural Network Model for Emotion Cause Analysis with Emotional Context Awareness. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4752–4757, Brussels, Belgium. Association for Computational Linguistics.

[3] Mathur, S. V., Jindal, A. R., Mittal, H., & Shrivastava, M. (2024). LastResort at SemEval-2024 Task 3: Exploring Multimodal Emotion Cause Pair Extraction as Sequence Labeling Task. *arXiv preprint arXiv:2404.02088*.

[4] Wang, F., Ding, Z., Xia, R., Li, Z., & Yu, J. (2022). Multimodal emotion-cause pair extraction in conversations. *IEEE Transactions on Affective Computing*.

[5] Nguyen, H. H., & Nguyen, M. T. (2023). Emotion-cause pair extraction as question answering. *arXiv preprint arXiv:2301.01982*.

[6] Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.