A Comparison Analysis Of Emotion-Cause Pair Extraction In Conversations

Prudhvi Nikku
snikku@umass.edu

Mani Kishan Ghantasala
mghantasala@umass.edu

Satya Sriram Potluri
spotluri@umass.edu

Srimathi Mahalingam
srimathimaha@umass.edu

Muskan Dhar
mdhar@umass.edu

1 Problem Statement

The goal of our project is to develop and analyze methods for Emotion-Cause Pair Extraction (ECPE) in conversations. This task involves not only identifying the emotions present in textual data but also uncovering the underlying causes of these emotions. From this study we aim to explore and compare few of many approaches proposed or implemented for this problem.

The problem statement is part of SemEval-2024 Task 3: Multimodal Emotion Cause Analysis in Conversations, a prestigious competition that challenges participants to develop models for extracting emotion-cause pairs from conversational data. Specifically, for the context of this study we are focusing on Subtask 1: Textual Emotion-Cause Pair Extraction in Conversations. This subtask requires us to identify both the emotions and their causes from text-based conversations, thereby providing a deeper understanding of the emotional dynamics within dialogues.

1.1 Motivations

- **Enhanced Emotional Intelligence in AI**: Current AI systems often lack the ability to understand and respond to human emotions effectively. By extracting emotion-cause pairs, AI can gain a deeper understanding of emotional expressions, leading to more empathetic and context-aware interactions. This is crucial for applications such as customer service chatbots, mental health assessment tools, and virtual assistants.

- **Applications in Various Domains**: ECPE has vast practical applications. In social media analysis, it can provide insights into public sentiment towards events or products, informing marketing and communication strategies. In mental health services, it can help therapists identify patterns in patient speech that indicate specific emotional triggers. In customer service, understanding the causes behind customer emotions can lead to more effective and personalized responses.

- **Challenges of Conversational Data**: Conversations are dynamic, with emotions and their causes intertwined within the flow of dialogue, influenced by context, tone, and the interplay between participants. Extracting emotion-cause pairs from such data presents unique challenges, requiring sophisticated models that can understand the subtleties of human communication.

- **Popularity and Limitations of ECE**: Emotion Cause Extraction (ECE) is a popular task in NLP, focusing on identifying causes of emotions in text. However, ECE alone often falls short in understanding the full context of conversations, as it typically deals with single-turn interactions. ECPE, on the other hand, introduces a more comprehensive understanding by addressing multi-turn dialogues and capturing the flow of conversation, leading to new insights and advancements in NLP tasks.

- **Advancing NLP Understanding and Learning**: ECPE introduces new dimensions of understanding and learning in NLP tasks. By focusing on the interplay between emotions and their causes in conversational contexts, ECPE can improve the ability of AI systems to process and interpret complex human interactions. This advancement is essential for creating more sophisticated and human-like AI systems that can better serve various practical applications.

1.2 Objectives

- **Developing Advanced Models for ECPE**: We aim to leverage state-of-the-art NLP models, including BERT, RoBERTa, DeBERTa, and Facebook's OPT 350M, to develop robust methods for emotion and cause extraction from conversational data.

- **Comprehensive Evaluation**: By implementing and comparing multiple models, we seek to identify the strengths and limitations of each approach. This involves evaluating the models using metrics such as accuracy, precision, recall, and F1-score, and analyzing their performance on a dataset specifically designed for ECPE in conversations.

- **Comparative Study**: Our project includes a comparative study of different approaches to ECPE. We will implement various methods and systematically compare their effectiveness in addressing the task. This will help in understanding the most effective strategies for emotion-cause pair extraction.

- **End-to-End Solution Using BiLSTM**: In addition to using advanced transformer-based models, we have developed an end-to-end solution using a BiLSTM network to explore its effectiveness in ECPE tasks.

- **Contributing to the Field of Emotion Analysis**: Our project aims to enhance the understanding of human emotions by focusing on the task of ECPE. We hope to provide valuable insights and methodologies that can be applied to various applications requiring emotional intelligence, contributing to ongoing research in NLP and emotion analysis.

The significance of our project lies in its potential to revolutionize the way AI systems understand and interact with human emotions. By extracting and analyzing emotion-cause pairs in conversations, we aim to develop AI that can engage with users in more meaningful and empathetic ways, ultimately enhancing the user experience across various domains.

2 What You Proposed vs. What You Accomplished

In this section, we summarize the tasks we proposed to complete in our project proposal and assess whether we accomplished them. Additionally, we provide brief explanations for any tasks that were not completed and note any significant changes to our project.

- ~~Develop an advanced model for Emotion-Cause Pair Extraction (ECPE)~~

- ~~Preprocess the ECF Dataset and create a custom dataset class~~

- ~~Train models on emotion detection and cause identification tasks~~

- ~~Integrate emotion detection and cause identification into a unified framework~~: Partially completed due to time constraints. We performed separate evaluations for each component.

- ~~Evaluate the models using metrics such as accuracy, precision, recall, and F1-score~~

- ~~Create visualizations of training metrics and results~~

- ~~Implement and evaluate a question-answering model for ECPE~~: Successfully introduced this paradigm and evaluated its effectiveness.

- *[New Introduction], We explored Promptuning to PEFT the LM for the ECPE task:* We encountered some challenges due to resource constraints, which limited our ability to train multiple models. Additionally, we are in the process of identifying the most effective approach to evaluate the outputs of these CasualLM versions of the models. **Reason**: We initiated this task a little later, and due to time constraints we are continuing to refine our evaluation methodology.

- ~~Prepare a comprehensive final report~~

Overall, we successfully accomplished most of the tasks we proposed in our project. The integration of emotion detection and cause identification into a unified framework was partially completed due to time constraints, but we managed to perform separate evaluations for each component and gained valuable insights into the effectiveness of our models. Additionally, we introduced and evaluated a question-answering paradigm for ECPE, which was not initially mentioned in our proposal but proved to be a valuable addition.

3 Related Work

The task of Emotion-Cause Pair Extraction (ECPE) has gained significant attention in recent years, reflecting the broader interest in understanding the causes behind emotions expressed in textual data. Khunteta and Singh (2021) (8) provide a comprehensive review of various methods and corpora used for emotion cause extraction, highlighting the evolution of techniques in this area. This section surveys prior work related to ECPE and highlights the evolution of methods used to tackle this challenging problem.

3.1 Early Approaches

Initial efforts in emotion cause extraction focused on rule-based and machine learning methods. For example, Lee et al. (9) introduced a rule-based system to detect the causes of emotions in text. Similarly, Neviarouskaya and Aono (**?**) employed a rule-based approach to extract emotion causes from text, demonstrating the potential of predefined rules in identifying causal relationships. Li and Xu (2014) (10) used text-based emotion classification to improve the extraction of emotion causes, demonstrating early integration of emotion and cause analysis.

3.2 Machine Learning and Statistical Models

With advancements in machine learning, researchers began to explore statistical methods for emotion cause extraction. Gao et al. (4) proposed the EC2C model to detect emotion causes in Chinese microblogs using machine learning techniques. Ghazi et al. (5) applied machine learning to detect emotion stimuli in emotion-bearing sentences, highlighting the potential of data-driven approaches.

3.3 Deep Learning Approaches

The advent of deep learning revolutionized the field of emotion cause extraction. Gui et al. (6) introduced a question-answering approach for emotion cause extraction using neural networks. Li et al. (11) developed a co-attention neural network model that incorporated emotional context awareness to analyze emotion causes, marking a significant advancement in the complexity and depth of the models used.

Xia and Ding (**?**) introduced the task of ECPE, which simultaneously extracts emotions and their causes, emphasizing the intricate relationship between emotional expressions and their triggers. Their work laid the foundation for subsequent research in this area.

3.4 Recent Advances

Recent studies have focused on leveraging advanced neural architectures to improve ECPE performance. Ding et al. (3) proposed integrating relative position and global label information to enhance emotion cause identification. Chen et al. (2) introduced the Recurrent Synchronization Network for ECPE, demonstrating state-of-the-art performance on benchmark datasets.

Singh et al. (**?**) presented an end-to-end network for ECPE, emphasizing the importance of combined training and layered architectures to capture the complex relationship between emotions and causes. Inoue et al. (7) proposed framing ECPE within a question-answering paradigm, highlighting innovative approaches to address the task.

3.5 Multimodal Approaches

The integration of multimodal data has also been explored in emotion analysis. McKeown et al. (**?**) and Busso et al. (1) investigated the use of multimodal records to understand emotional expressions better. Wang et al. (**?**) specifically constructed a multimodal dataset for ECPE in conversations, illustrating the potential of combining textual, auditory, and visual data.

4 Dataset

For our project, we use the Emotion-Cause-in-Friends (ECF) dataset (**?**), which is specifically constructed for the task of Multimodal Emotion-Cause Pair Extraction in Conversations. This dataset is derived from the popular American sitcom *Friends*, known for its rich emotional content and diverse conversational contexts.

4.1 Dataset Properties and Statistics

The ECF dataset contains the following properties and statistics:

- **Conversations**: The dataset includes a total of 1,344 conversations, each composed of multiple utterances, forming dialogues between two or more speakers.

- **Total Utterances**: There are 13,509 individual utterances within the dataset, each representing a single piece of spoken or written communication within a conversation.

- **Annotated Utterances with Emotions**: Out of the total utterances, 7,528 have been annotated with specific emotions, identifying the emotional context of each utterance.
- **Emotion-Cause Pairs**: The dataset contains 9,272 emotion-cause pairs, each consisting of an identified emotion and its corresponding cause within the conversation, providing a deeper understanding of why a particular emotion was expressed.
- **Emotions**: The dataset focuses on six basic emotions, each represented by a significant number of pairs:

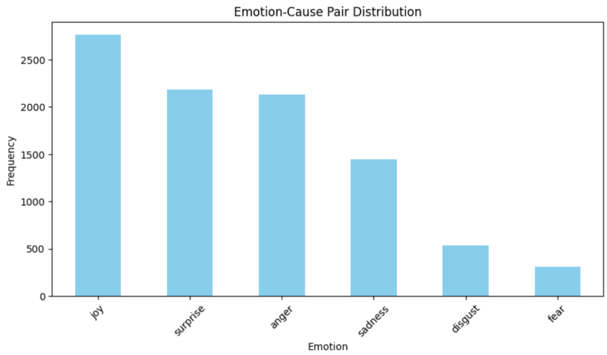

Figure 1: Distribution of Emotions in the ECF Dataset

As shown in Figure 1, the distribution of the six basic emotions in the ECF dataset. Each emotion is well-represented, providing a diverse set of examples for model training and evaluation. This distribution helps in understanding the prevalence of each emotion in the dataset and ensures that the models are trained on a balanced set of emotional expressions.

4.1.1 Conversation Length Observations

We analyzed the length of conversations in the ECF dataset to understand their structure. The average conversation length is 10 utterances, with a standard deviation of 5. Longer conversations tend to have more complex emotional dynamics and multiple emotion-cause pairs, making them more challenging for models to analyze.

4.1.2 Average Sentiment per Emotion

The average sentiment scores for each emotion were calculated using the VADER sentiment analyzer. Positive emotions such as joy have higher average sentiment scores, while negative emotions like sadness and anger have lower scores. This helps in understanding the sentiment distribution within the dataset and provides insights into how different emotions are expressed in conversations.

4.2 Challenges

The dataset presents several challenges:

- **Conversational Dynamics**: Emotions and their causes are intertwined within the flow of dialogue, influenced by context, tone, and participant interplay.
- **Implicit Causes**: Emotional triggers are often implied rather than explicitly stated, requiring models to understand nuanced conversational cues.
- **Contextual Relevance**: Extracting relevant causes from a large amount of conversational data demands sophisticated models capable of discerning contextually significant information.

4.3 Example Input/Output Pairs

To illustrate our task, we present a couple of input/output pairs from the ECF dataset:

```
"conversation_ID": 3,
"conversation": [
{
    "utterance_ID": 1,
    "text": "I do not want to
    be single, okay?
    I just... I just... I just
    wanna be married again!",
    "speaker": "Ross",
    "emotion": "sadness"
},
{
    "utterance_ID": 2,
    "text": "And I just want
    a million dollars!",
    "speaker": "Chandler",
    "emotion": "neutral"
},
{
    "utterance_ID": 3,
    "text": "Rachel?!",
    "speaker": "Monica",
    "emotion": "surprise"
}
],
"emotion-cause_pairs": [
[
    "1_sadness",
    "1_I do not want to
    be single"
],
[
    "3_surprise",
    "3_Rachel?!"
]
]
```

These examples highlight the nature of the data and the extraction task. The dataset's diversity and complexity require our models to effectively capture and understand both explicit and implicit emotional triggers within conversational contexts.

4.4 Data Preprocessing

For preprocessing, we performed the following steps:

- **Tokenization**: Used the BERT tokenizer to tokenize the conversational data.
- **Lowercasing**: Converted all text to lowercase to maintain consistency.
- **Padding and Truncation**: Applied padding and truncation to ensure uniform input lengths for the model.
- **Special Tokens**: Added special tokens to delineate different parts of the conversation.

These preprocessing steps were essential to prepare the data for input into our models, ensuring that the text was in a suitable format for tokenization and model training.

4.5 Data Annotation

Our project did not involve new data annotation but relied on the annotations provided in the ECF dataset. The dataset annotations include emotion labels and the corresponding causes within the conversational context. We focused on utilizing these existing annotations to train and evaluate our models.

5 Baseline

Baseline methods serve as essential reference points in the evaluation of sophisticated models. For the task of Emotion-Cause Pair Extraction (ECPE), establishing robust baselines allows for meaningful comparisons and highlights the strengths and weaknesses of more complex approaches. In this section, we detail the baseline methods employed in our study, focusing on their design, implementation, and evaluation. We explored both rule-based and machine learning-based methods as our baselines. The rule-based method, which relies on predefined keywords and patterns, demonstrated significant limitations in capturing the nuances of natural language, resulting in poor performance. In contrast, the machine learning-based method, utilizing logistic regression, showed more promise by leveraging statistical patterns in the data, providing a stronger foundation for comparison with more advanced models.

5.1 Rule-Based Model

Working Principle: The rule-based model classifies emotions and identifies causes using predefined keywords and patterns. Emotions are assigned based on the presence of specific keywords within the text, while causes are identified through the detection of causal conjunctions such as ”because,” ”due to,” and ”as a result of.”

Emotion Classification Rules: The rule-based model uses specific keywords to identify different emotions within the text. Table 1 lists the keywords used for each emotion.

Emotion	Keywords
Joy	happy, joyful, delighted
Anger	angry, furious, mad
Sadness	sad, unhappy, sorrowful
Surprise	surprised, shocked, amazed
Neutral	Default
Fear	scared, afraid, terrified
Disgust	disgusted, repulsed, horrified

Table 1: Emotion Classification Rules

Cause Identification Rules: The model looks for causal indicators such as ”due to,” ”because,” ”as a result,” ”since,” and ”caused by” to identify potential causes within the text.

Results:

Overall Accuracy for Emotion Classification: 0.44

Overall Accuracy for Cause Classification: 1.00

Inefficiency of Rule-Based Model: While

Emotion	Precision	Recall	F1-Score
Joy	0.52	0.02	0.04
Surprise	0.00	0.00	0.00
Anger	0.14	0.01	0.02
Sadness	0.23	0.03	0.05
Disgust	0.00	0.00	0.00
Fear	0.00	0.00	0.00
Neutral	0.45	0.98	0.62

Table 2: Emotion Classification Results

Cause Class	Precision	Recall	F1-Score
Not Cause (0)	1.00	1.00	1.00
Cause (1)	1.00	1.00	1.00

Table 3: Cause Classification Results

the rule-based model is straightforward and interpretable, it suffers from several limitations that make it inefficient for emotion-cause pair extraction:

- **Keyword Dependency:** The model relies heavily on predefined keywords, which limits its ability to generalize across different contexts and variations in language.

- **Keyword Dependency:** The model relies heavily on predefined keywords, which limits its ability to generalize across different contexts and variations in language.

- **Lack of Context Understanding:** The rule-based approach cannot capture the nuances and complexities of natural language, often leading to misclassifications.

- **Low Performance Metrics:** As seen in the results, the model exhibits low precision, recall, and F1-scores for most emotions, indicating poor performance and reliability.

- **Binary Cause Identification:** The model's perfect scores in cause classification suggest it is not effectively discerning between causes and non-causes but rather identifying them based on simple keyword presence.

Due to these inefficiencies, the rule-based model serves as a foundational benchmark but falls short in accurately and robustly extracting emotion-cause pairs from text, underscoring the need for more advanced machine learning approaches.

5.2 Machine Learning-Based Model

Working Principle: This model utilizes logistic regression to classify emotion-cause pairs. Logistic regression was chosen due to its simplicity and interpretability, providing a clear understanding of the relationships between features and the target variable.

Feature Extraction: We employed term frequency-inverse document frequency (TF-IDF) scores and n-grams to numerically represent the text data. These features capture the importance of words and their combinations in the context of the given corpus, enabling the model to identify patterns and relationships relevant to emotion-cause pair extraction.

Hyperparameter Tuning: We conducted grid search and cross-validation to optimize hyperparameters, including:

- **Regularization Strength (C):** Tested a range of values to balance the trade-off between fitting the training data and generalizing to unseen data.

- **Solver Type:** Evaluated different solvers such as 'liblinear', 'sag', and 'lbfgs' to determine the most effective optimization algorithm for logistic regression.

Train/Validation Split: To ensure robust evaluation, we split our dataset into training and validation sets as follows:

- **Training Set:** 80% of the data

- **Validation Set:** 20% of the data

This split allowed us to tune hyperparameters on the validation set and evaluate the final model performance, ensuring that the results are generalizable and not overfitted to the training data.

Results:
Overall Accuracy for Emotion Classification: 53%
Overall Accuracy for Cause Classification: 65%

Rationale for Selection: Logistic regression provides a more nuanced approach than rule-based models by leveraging statistical patterns in the data. This model serves as a step up from rule-based methods, demonstrating the potential of machine learning techniques in capturing more complex relationships within the data.

Emotion	Precision	Recall	F1-Score
Joy	0.56	0.31	0.40
Surprise	0.61	0.46	0.53
Anger	0.37	0.19	0.25
Sadness	0.36	0.12	0.18
Disgust	0.00	0.00	0.00
Fear	0.00	0.00	0.00
Neutral	0.54	0.88	0.67

Table 4: Emotion Classification Results

Cause Class	Precision	Recall	F1-Score
Not Cause (0)	0.67	0.66	0.67
Cause (1)	0.64	0.65	0.65

Table 5: Cause Classification Results

Our baseline study provided valuable insights into the strengths and limitations of both rule-based and machine learning approaches for the ECPE task. While the rule-based model served as a foundational benchmark, its inefficiencies underscored the need for more advanced techniques. The logistic regression model highlighted the potential for significant performance improvements with more sophisticated methods. This comprehensive evaluation guided the development and refinement of our End-to-End Network for Emotion-Cause Pair Extraction, ultimately leading to competitive performance in identifying emotion-cause pairs.

By establishing these baselines, we ensured a thorough and robust evaluation framework, facilitating the development of advanced models that significantly push the boundaries of ECPE performance.

6 Our Approach

In this study, we aim to perform a comparative analysis of various approaches to the Emotion-Cause Pair Extraction (ECPE) task. ECPE is a crucial task in natural language processing, particularly for understanding and analyzing conversational data. By comparing different models, we seek to identify the strengths and weaknesses of each approach and determine the most effective methods for joint emotion and cause extraction.

6.1 Model Architectures:

In this section, we explore the various architectures employed in our study for Emotion-Cause Pair Extraction (ECPE). Each model is designed to capture the intricate relationships between emotions and their causes within the text using different paradigms and architectures. We begin with an in-depth look at each architecture and then discuss other approaches evaluated in our comparative analysis.

6.1.1 End-to-End architecture (E2E $PExt_E$):

The first approach we examined is the End-to-End emotion cause pair extraction model (E2E $PExt_E$) by Aaditya Singh et al. (**?**). This model processes an entire document to compute, for each ordered pair of clauses (c_i, c_j), the probability of being a potential emotion-cause pair.To enhance the learning of suitable clause representations for this primary task, the model is also trained on two auxiliary tasks: Emotion Detection and Cause Detection.

While this was built on the dataset introduced by (Fan et al., 2019; Li et al., 2019) (**?**) for Emotion-Cause Extraction (ECE), we inspired from this approach and adapted the architecture to work on the selected dataset for the study.

This model architecture consists of several key components, following a hierarchical structure proposed by Singh et al. (**?**):

1. **Word-Level Encoder (BiLSTM + Attention)**: This component takes the words of each clause as input and generates word-level embeddings using a BiLSTM with an attention mechanism. These embeddings capture the contextual meaning of each word within its clause.

2. **Clause-Level Encoder (BiLSTM)**: The word-level embeddings are further processed by another BiLSTM to generate clause-level representations. These contextualized clause representations are then used for the classification tasks.

3. **Auxiliary Tasks (Emotion and Cause Detection)**: To facilitate learning suitable clause representations, the model is also trained on two auxiliary tasks: Emotion Detection and Cause Detection. This interactive approach improves performance on the primary ECPE task, as observed by Xia and Ding (2019).

4. **Pair Predictor (Fully Connected Network)**: The emotion and cause clause embeddings are combined using a Cartesian

product to form potential emotion-cause pairs. These pairs are then passed through a fully connected network to predict the likelihood of each pair being a valid emotion-cause pair.

The intuition and motivation for this hierarchical architecture come from the observation that performance on auxiliary tasks can be improved when done interactively rather than independently.

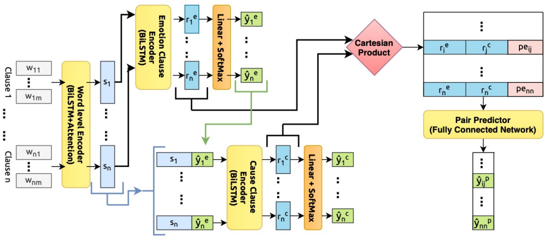

Figure 2: End-to-End Network for Emotion-Cause Pair Extraction (E2E-PExtE)

As illustrated in the Figure 2. The input consists of multiple clauses, each represented by a series of words. A BiLSTM with an attention mechanism encodes the words into embeddings, which are then processed by additional BiLSTMs to generate emotion and cause clause embeddings. These embeddings are used to predict the probability of each clause containing an emotion or cause. The emotion and cause embeddings are combined using a Cartesian product to form potential emotion-cause pairs, which are then processed by a fully connected network to predict the likelihood of each pair being valid.

6.1.2 EmoBERTa-Based Approach

The next approach we utilized for the Emotion-Cause Pair Extraction (ECPE) task is based on the EmoBERTa model introduced by (Taewoon et al) (**?**), which is designed for speaker-aware emotion recognition in conversation using the RoBERTa architecture. This model leverages transformer architecture for accurate emotion classification and cause identification in dialogues.

EmoBERTa for Emotion Recognition We initialized the EmoBERTa model for sequence classification and fine-tuned it for the emotion classification task. This model classifies each utterance in a dialogue into one of the predefined emotion classes found in the dataset. The fine-tuning was performed focusing on correctly classifying the emotion of the utterances.

Transformer Encoder for Emotion-Cause Pair Extraction Following the emotion classification, we employed a Transformer Encoder model fine-tuned for the emotion-cause pair extraction task. The goal of this model is to identify the utterance in a given conversation that may be the cause of the current utterance's emotion, not specifically dependent on the emotion but the sentence itself.

Emotion-Cause Mapping The mapping logic combines the outputs from the emotion classification and the cause pair extraction models. Specifically, it pairs the identified emotions with their corresponding cause utterances when the detected emotion is not neutral.

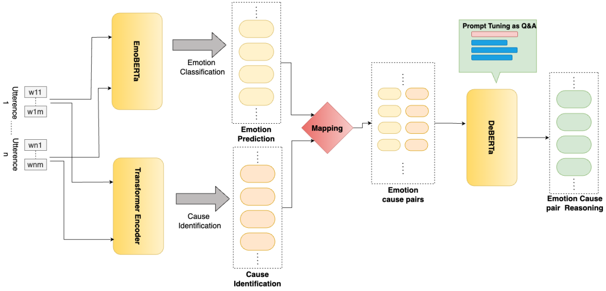

Figure 3: Architecture of the EmoBERTa-Based Approach

Extended Pair Extraction with DeBERTa To enhance the pair extraction process, we incorporated another model DeBERTa which is designed for question answering tasks. This model, trained on question-answer pairs including unanswerable questions, was prompt-tuned with the emotion-cause pairs extracted from the mapping function. The purpose of this tuning was to refine the identification of the dialogue segment that serves as the cause of the detected emotion.

6.1.3 Prompt Tuning Based Approach

The final approach we utilized for the Emotion-Cause Pair Extraction (ECPE) task involves prompt tuning, leveraging a pre-trained language model for a specific sequence classification task. This method enhances the model's ability to extract emotion-cause pairs from conversational data by fine-tuning it with custom prompts.

Pre-trained Model For this approach, we used the pre-trained language model "facebook/opt-350m" from Hugging Face's transformer library. OPT was first introduced in Open Pre-trained Transformer Language Models and first released in metaseq's repository on May 3rd 2022 by Meta AI. OPT was predominantly pretrained with English text, but a small amount of non-English data is still present within the training corpus via CommonCrawl. The model was pretrained using a causal language modeling (CLM) objective. OPT belongs to the same family of decoder-only models like GPT-3. As such, it was pretrained using the self-supervised causal language modedling objective.

Formatting Prompts The data was formatted into prompts that included the conversational context and specific instructions to identify emotions and their causes. Each prompt consisted of the context, the target emotion, and the cause utterance, structured to guide the model in recognizing these elements.

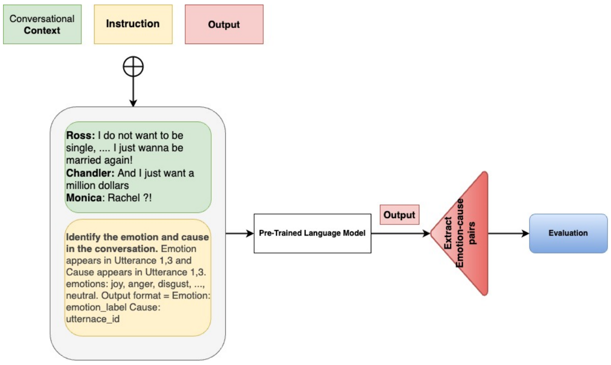
Conversational Context
Instruction
Output
Ross: I do not want to be single, ... I just wanna be married again!
Chandler: And I just want a million dollars
Monica: Rachel ?!
Identify the emotion and cause in the conversation. Emotion appears in Utterance 1,3 and Cause appears in Utterance 1,3. emotions: joy, anger, disgust, ..., neutral. Output format = Emotion: emotion_label Cause: utternace_id
Pre-Trained Language Model
Output
Extract Emotion-cause pairs
Evaluation

Figure 4: Architecture used for Prompt-Tuning LMs

Fine-Tuning Process The fine-tuning process involved setting up training parameters (batch size, learning rate, gradient accumulation steps), using a specialized framework to train the pre-trained model on custom prompts over multiple epochs, and continuously adapting the model to better understand the task of emotion-cause pair extraction.

6.2 Implementation Details

Computational Resources For the implementation and experimentation of our models, we utilized the following computational resources:

Google Colab (Free Version):

- **CPU**: Intel Xeon CPU with 2 vCPUs
- **GPU**: 1 x NVIDIA Tesla T4 16GB VRAM
- **RAM**: Up to 12.6 GB.

Kaggle (Free Version):

- **CPU**: Intel Xeon CPU with 2 vCPUs
- **GPU**: 2 x NVIDIA Tesla T4 16GB VRAM
- **RAM**: Up to 29 GB.

6.2.1 End-to-End architecture (E2E $PExt_E$)

Completion status: Fully implemented and evaluated.

Implementation:

- The hierarchical architecture was implemented using BiLSTM layers for both word-level and clause-level encoding.
- Auxiliary tasks of Emotion Detection and Cause Detection were incorporated to enhance the main task.
- Cartesian product used to form potential emotion-cause pairs, followed by a fully connected network for classification.

Tricks implementing: There were a lot of issues with the w2v interpretations about unknown tokens in evaluation data as tokenization was done with out any AutoTokenizers, After alot of tweaking we managed to handle the tokenization errors. Out of those the major tweaks are handling UNK tokens, and tokenizing a bigger corpa and adding the words from the both datasets.

6.2.2 EmoBERTa-Based Approach

Completion status: Fully implemented and evaluated.

Implementation:

- EmoBERTa model was specifically used for capturing emotional contexts in conversations. It's initialized with configurations suited for sequence classification and adapted to identify various emotions based on training data.
- A Transformer Encoder for Cause Identification consists of layers designed to process EmoBERTa outputs, pinpointing cause utterances linked to identified emotions by utilizing contextual embeddings to enhance cause identification accuracy.
- Mapping logic created to pair emotions with their causes, with additional refinement using a DeBERTa model for precise emotion-cause

pair extraction using Parameter Efficient Fine Tuning.

- **Parameter-Efficient Fine-Tuning:** The DeBERTa model uses parameter-efficient fine-tuning to enhance specific capabilities with minimal updates. This method achieves effective updates with a reduced computational load.

- **Leveraging a Question-Answering Framework:** The system adopts a question-answering approach to identify and correlate emotions with their causes within dialogues. This framework mimics human reasoning, allowing for precise extraction of conversational dynamics.

- **Example Prompt**: **Prompt PlaceHolder:** Which part of the text '{}' is the reason for "{}" 's feeling of "{}" when "{}" is said?

  Prompt Example: "Which part of the text 'I do not know . We are talking about whipped fish , Monica . I am just happy I am keeping it down , you know ?' is the reason for ' Joey "s feeling of ' disgust ' when ' I do not know . We are talking about whipped fish , Monica . I am just happy I am keeping it down , you know ? ' is said?"

- **LoRA Adaptation:** LoRA adaptation is implemented to efficiently fine-tune large pre-trained models by modifying only a small subset of parameters. This approach enhances memory efficiency, crucial in scenarios with limited computational resources.

Tricks and blockers while implementing:

- **Tricks**: Resource-Efficient Adaptation: By employing LoRA, the implementation effectively manages resource utilization, allowing for the adaptation of robust models within a constrained resource environment. This approach not only preserves the pre-trained models' capabilities but also ensures that the system remains scalable and efficient.

6.2.3 Prompt Tuning-Based Approach

Completion status: Partially implemented and stuck at finding reliable evaluation techniques.

Implementation:

- Formatted data into prompts with clear instructions for identifying emotions and causes.

- Fine-tuned the model using these prompts, adapting it to the task of emotion-cause pair extraction.

- Explored various LLMs from Hugging face that can be trained by the resources we have. Some models we explored are Vicuna, LLama2, OPT-350M.

- we tried to use Quantized models to make the models lighter but that didn't work for all the models explored.

Blockers while implementing:

- Formatting prompts in a way that consistently yielded accurate emotion and cause identification; managing the fine-tuning process to avoid overfitting.

- Not able to get smaller LLM like OPT-350M to generate the output in the desired output format which helps to evaluate model didn't work. It generated random or repeated words for 40% for cases.

- The compute time limitations and GPU memory limitations didn't allow to use larger models or run multiple experiments within the timeframe of the project.

6.3 Results and Comparison

In this section we will be comparing the results of the performance of different approaches we explored in the study. We use the same evaluation metrics (precision, recall and F1-score), as used in the past work on ECE and ECPE tasks. the metric definitions are defined as:

$$P = \frac{\#correct\_pairs}{\#proposed\_pairs} \tag{1}$$

$$R = \frac{\#correct\_pairs}{\#annotated\_pairs} \tag{2}$$

$$F1 = \frac{2 * P * R}{P + R} \tag{3}$$

where,

- $\#proposed\_pairs$ = no. of emotion-cause pairs predicted by the model

- $\#correct\_pairs =$ number of emotion-cause pairs predicted correctly by the model
- $\#annotated\_pairs =$ total number of actual emotion-cause pairs in the data

P, R, and $F1$ for the two auxiliary classification tasks (emotion-detection and cause-detection) have the usual definition.

We have used these three metrics to compare the performance of emotion-cause pair extraction along with the intermediary tasks such as Emotion Identification and Cause Identification across all applicable architectures/models. This comprehensive evaluation helps in understanding not only the effectiveness of the models in extracting emotion-cause pairs but also their accuracy in identifying individual emotions and causes.

The comparison involves:

- **End-to-End architecture (E2E $PExt_E$)**: Evaluated for its performance in directly predicting emotion-cause pairs from the input data using hierarchical BiLSTM layers. Additionally logged the metrics for Emotion Classification and Cause Identification tasks too.
- **EmoBERTa-Based Approach**: Assessed for its ability to classify emotions in conversations using the EmoBERTa model and subsequently identify causes with the Transformer Encoder model.
- **Prompt Tuning-Based Approach**: Unfortunately we weren't able to come up with a reliable evaluation method that extracts the required fields from the outputs of LLMs, this is coupled with the resource constraints which is why we could only test the prompt tuning on smaller models (300-400M parameters) like opt-350m which is not always generating output in the desired output format for easy rule based extraction. There are some instances the fine-tuned model generating random texts.

The table 6, shows the performance metrics of all these methods.

7 Error analysis

We have collected set of 112 examples where the E2E-PEXT$_E$ and EmoBERTa models have fails i.e, the emotion-cause pairs or emotion detection were incorrectly predicted. Upon analyzing these cases, we observed the following errors:

1. Misclassification of emotions.
2. Incorrect pair identification.

Observations

- **Exclamation Marks:** The models tend to predict high confidence for surprise when exclamation marks are present.
- **Sentence Length:** Longer sentences pose a challenge for the models in accurately identifying the correct pairs.
- **Simple Sentences:** Instances such as ”Okay, Okay” have been incorrectly identified as causes, which does not align with manual reading. This may indicate that the models have detected a pattern in the dataset.
- **Missing Pairs:** There are cases where no pairs were found, possibly due to the elimination of sentences being their own causes. This might be an error, and further investigation is required.

These observations highlight areas where the models can be improved to enhance their accuracy and reliability in emotion-cause pair extraction.

When we fine-tuned OPT-350m, we were unable to find evaluation logic to extract output correctly. Some outputs appeared in the following manner:

```
__ expected_ expected
expected_ expected__ expected___ West__
we__ expected_____ town___?

expected ### expected expected
```

We require more time to debug or fix these issues, so we do not have any reportable results at this time.

8 Contributions

This project was a collaborative effort with each member contributing significantly to different aspects. The detailed contributions are as follows:

- **Prudhvi Nikku:** Data collection, processing, and DeBERTa - EmoBERTa.
    - Collected and preprocessed the Emotion-Cause-in-Friends (ECF) dataset.

Table 6: Experiment results on emotion, cause, and pair extraction tasks on the test datasets.

	Emotion Extraction			Cause Extraction			Pair Extraction		
Method	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score
Baseline	0.4900	0.5400	0.4800	0.6600	0.6600	0.6600	0.4574	0.2213	0.2983
E2E-PExt$_E$	0.7423	0.7188	0.7303	0.6841	0.6115	0.6458	0.4848	0.6267	0.5467
EmoBERTa	0.5566	0.5272	0.5100	0.0230	0.0300	0.0270	0.2416	0.2399	0.2312

  - Conducted initial data analysis, including creating visualizations for emotion distribution and conversation length.
  - Implemented DeBERTa and EmoBERTa based architecture for ECPE task.
  - Integrated EmoBERTa for parameter-efficient fine-tuning with quantized LoRA.

- **Mani Kishan Ghantasala:** E2E PEXT Model development and fine-tuning.
  - Developed and fine-tuned an end-to-end emotion cause extraction model E2E-PEXT.
  - Developed and experimented prompt tuning on Illama2 4-bit quantized model.
  - Defined and structured the Prompt tuning techniques including the prompt formats, and other logics.
  - Created detailed architecture diagrams illustrating model workflows.

- **Srimathi Mahalingam:** Prompt-based tuning and report compilation.
  - Implemented a Transformer Encoder model for Cause Identification.
  - Evaluated the performance of the Vicuna, llama2 with prompt-tuning techniques.
  - Compiled the final report, ensuring all sections were well-integrated and cohesive.
  - Analyzed the results of prompt tuning and contributed to refining the models.
  - Assisted in the documentation of model refinement processes and results.

- **Muskan Dhar:** Model analysis and refinement, baseline models.
  - Performed extensive error analysis to identify and address model weaknesses.
  - Evaluated model performance using various metrics and provided insights for improvement.
  - Implemented Prompt tuning technique on OPT-350M for ECPE.
  - Implemented baseline models for emotion and cause classification using logistic regression.

- **Satya Sriram Potluri:** Prompt tuning and model experimentation.
  - Integrated EmoBERTa for parameter-efficient fine-tuning with quantized LoRA.
  - Led the prompt tuning efforts using various models.
  - Designed and implemented experiments to compare the effectiveness of different prompt-tuning strategies.
  - Explored 4-bit quantized models for prompt-based tuning.

9 Conclusion

This study investigated several methods for Emotion-Cause Pair Extraction (ECPE) in conversations, including an end-to-end architecture, an EmoBERTa-based approach, and prompt tuning with pre-trained language models. Our evaluation demonstrated the potential of Parameter Efficient Fine-Tuning (PEFT) as a promising direction for enhancing ECPE performance. Leveraging advanced language models such as GPT-3.5, Mistral, and LLama, in conjunction with techniques like Flash Attention and DeepSpeed, can significantly accelerate the training process and optimize resource usage.

Given our current focus on textual data, we recognize that incorporating multimodal data—such as video and audio—alongside text could provide a more comprehensive understanding of conversational context and yield better metrics. Future work will involve exploring these multimodal approaches to further improve ECPE performance and deepen our understanding of emotion-cause dynamics in conversations.

10 AI Disclosure

- Did you use any AI assistance to complete this proposal? If so, please also specify what AI you used.
  - Yes, ChatGPT.

If you answered yes to the above question, please complete the following as well:

- If you used a large language model to assist you, please paste *all* of the prompts that you used below. Add a separate bullet for each prompt, and specify which part of the proposal is associated with which prompt.
  - Prompt for Introduction and Problem Statement:
    - ∗ ”Provide a detailed problem statement for an Emotion-Cause Pair Extraction (ECPE) project, including motivations and objectives.”
  - Prompt for Related Work Section:
    - ∗ ”Summarize the recent advances in emotion-cause pair extraction research.”
  - Prompt for Baseline Models:
    - ∗ ”Describe the rule-based and machine learning-based baseline models for emotion-cause pair extraction.”
  - Prompt for Our Approach Section:
    - ∗ ”Explain the End-to-End architecture and the EmoBERTa-based approach for ECPE.”
  - Check for any grammatical or spelling errors in this.
- **Free response:** For each section or paragraph for which you used assistance, describe your overall experience with the AI. How helpful was it? Did it just directly give you a good output, or did you have to edit it? Was its output ever obviously wrong or irrelevant? Did you use it to generate new text, check your own ideas, or rewrite text?
  - The AI was helpful in generating the initial drafts of few sections of the proposal. For the most part, it provided a solid foundation that only required minor edits and additions to align with our specific project details and goals. The outputs were generally relevant and well-structured, although some parts needed to be rephrased or expanded to fit the context better. The AI was used to check the coherence of our ideas, and rewrite sections for clarity and conciseness.

References

[1] Busso, C., Bulut, M., Lee, C.-C., Kazemzadeh, E. A., Provost, E. M., Kim, S., Chang, J. N., Lee, S., and Narayanan, S. S. (2008). Iemocap: Interactive emotional dyadic motion capture database. *Language Resources and Evaluation*, 42:335–359.

[2] Chen, F., Shi, Z., Yang, Z., and Huang, Y. (2022). Recurrent synchronization network for emotion-cause pair extraction. *Knowledge-Based Systems*, 238:107965.

[3] Ding, Z., He, H., Mengran, Z., and Xia, R. (2019). From independent prediction to reordered prediction: Integrating relative position and global label information to emotion cause identification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33:6343–6350.

[4] Gao, K., Xu, H., and Wang, J. (2015). Emotion cause detection for Chinese micro-blogs based on ecocc model. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, pages 3–14.

[5] Ghazi, D., Inkpen, D., and Szpakowicz, S. (2015). Detecting emotion stimuli in emotion-bearing sentences. In *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics*.

[6] Gui, L., Hu, J., He, Y., Xu, R., Lu, Q., and Du, J. (2017). A question answering approach for emotion cause extraction. In Palmer, M., Hwa, R., and Riedel, S., editors, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1593–1602, Copenhagen, Denmark. Association for Computational Linguistics.

[7] Inoue, S., Nguyen, M.-T., Mizokuchi, H., Nguyen, T.-A., Nguyen, H.-H., and Le, D. (2023). Towards safer operations: An expert-involved dataset of high-pressure gas incidents for preventing future failures. In Wang, M. and Zitouni, I., editors, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages

509–521, Singapore. Association for Computational Linguistics.

[8] Khunteta, A. and Singh, P. (2021). Emotion cause extraction - a review of various methods and corpora. In *Proceedings of the 2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC)*, pages 314–319.

[9] Lee, S. Y. M., Chen, Y., and Huang, C.-R. (2010). A text-driven rule-based system for emotion cause detection. In Inkpen, D. and Strapparava, C., editors, *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.

[10] Li, W. and Xu, H. (2014). Text-based emotion classification using emotion cause extraction. *Expert Systems with Applications*, 41:1742–1749.

[11] Li, X., Song, K., Feng, S., Wang, D., and Zhang, Y. (2018). A co-attention neural network model for emotion cause analysis with emotional context awareness. In Riloff,