Textual Emotion-Cause Pair Extraction in Conversations

Shruti Mandaokar, Suparshwa Patil and Shreyas Bedekar
mands01@pfw.edu, patis04@pfw.edu, bedesn01@pfw.edu
Purdue University Fort Wayne

Abstract

The accurate extraction and comprehension of emotion-cause pairs from textual data are pivotal for enhancing artificial intelligence systems' ability to interpret human emotions and their triggers. Traditional approaches in Emotion-Cause Extraction (ECE) often require prior emotion annotations, limiting their applicability in real-world scenarios and overlooking the interconnected nature of emotions and their causes. This research introduces a novel task, Emotion-Cause Pair Extraction (ECPE), which first identifies the emotion and associates emotional expressions with their causative contexts within a text document. We leverage state-of-the-art transformer-based models such as Logistic Regression with TFIDF for Emotion Classification and DistilBERT from Hugging Face for Cause Extraction and achieve an accuracy of 72.04% for ECPE on Friends Dataset from Semeval Task3.

Additionally, the paper talks about our experiments and results with BERT, RoBERTa, and DistilBERT from Hugging Face, combined with Logistic Regression and TFIDF techniques, to dissect and analyze conversational texts.

Keywords: Multimodal Emotion Analysis, Emotion-Cause Extraction, Logistic Regression, BERT, Tokenization, RoBERTa, Conversational Data

1 Introduction

Emotions significantly influence human communication dynamics, prompting the need for effective methods to understand and analyze them, particularly in conversational settings. Emotion-Cause Pair Extraction (ECPE) advances sentiment analysis by not only identifying emotions in text but also linking them to their specific causes. This dual extraction offers deeper insights into text data and is applied in more advanced contexts such as in-depth customer service, therapy, and conversational agents, where understanding the reason behind an emotion is crucial. Early foundational work by Xia and Ding (2019) defines the scope and challenges of ECPE, highlighting its importance for advanced AI systems that require an understanding of nuanced human emotions[1].

In the domain of Natural Language Processing (NLP), the novel task of Emotion-Cause Pair Extraction (ECPE) presents a multifaceted challenge that extends beyond traditional sentiment analysis. The fundamental complexity lies in the contextual understanding required to discern the intricate relationship between expressed emotions and their underlying causes in textual conversations. Unlike standard emotion detection tasks, ECPE must contend with several nuanced difficulties.

First, the identification of emotion is seldom straightforward due to the potential for multiple causes within a single conversational thread, necessitating the model to differentiate and prioritize causes with varying degrees of relevance. Second, temporal dependencies within dialogues introduce a layer of complexity, as the cause of an emotion may not be immediately preceding but could be inferred from earlier segments of the conversation.

Furthermore, the phenomenon of cross-utterance emotional causes challenges the model to maintain coherence over dispersed textual cues, which can be subtle and context-dependent. Lastly, the intricate task is compounded by linguistic elements such as idiomatic expressions, sarcasm, and cultural nuances that can obscure the emotion-cause relationship. Our research is driven by the aspiration to enhance emotionally intelligent systems, refining their ability to interpret the intricacies of human emotion and context within conversations.

Subtask 1 of the Multimodal Emotion Cause Analysis in Conversations (ECAC) of SemEval 2023 Task 3 centers on the extraction of emotion-cause pairs from textual dialogues. This study approaches the challenge by segmenting the task into two distinct phases. Initially, we implement Emotion Classification by deploying a Logistic Regression model in conjunction with TFIDF vectorization to systematically categorize the expressed emotions. Subsequently, the study progresses to the Cause Extraction phase, where we employ the RoBERTa model, re-purposed as a Question Answering system, to discern the underlying causes behind the identified emotions. This bifurcated methodology not only enhances the precision of emotion identification but also enriches the contextual understanding necessary for accurate cause detection.

2 Motivation

Traditional sentiment analysis and emotion detection techniques focus on identifying the polarity or intensity of emotions expressed in text. While valuable, these approaches lack granularity in understanding the underlying reasons or stimuli behind the expressed emotions. Emotion Cause Pair Extraction aims to bridge this gap by identifying the causal factors that lead to specific emotional states. By unraveling the causal relationships between language and emotions, we can enhance the interpretability and effectiveness of affect-based models, leading to more contextually relevant and empathetic AI systems. Emotion Cause Pair Extraction aims to fill this gap by identifying the causal relationships behind emotional expressions, enhancing the interpretability and functionality of emotion analysis systems[2]. As Chen et al. (2020) discuss, understanding these causal relationships can lead to more empathetic and contextually aware AI systems[3].

3 Problem Statement

The task involves extracting emotion-cause pairs from conversational data, where an emotion cause is defined and annotated as a textual span. Given a conversation containing speaker information and utterances, the goal is to output all emotion-cause pairs, each containing an emotion utterance along with its emotion category and the textual cause span in a specific cause utterance. This task is essential for subsequent steps in emotion analysis, such as emotion-cause pair extraction, sentiment analysis, and dialogue systems.

4 Objectives

Our main objectives in the work include:

- Develop a robust methodology for emotion analysis in conversational data to accurately extracting emotion-cause pairs.
- Explore transformer-based architectures such as BERT and RoBERTa for capturing contextual information.
- Investigate techniques for emotion classification, cause extraction, and emotion transition prediction within conversations.
- Evaluate the performance of the proposed approach using comprehensive metrics and comparative analyses.

5 Dataset Structure

The provided dataset consists of two JSON files: **Subtask_1_train.json** and **Subtask_1_test.json**. Each file contains conversations with associated utterances and emotion-cause pairs.

5.1 Train Data

Subtask_1_train.json contains conversations used for training the model. Each conversation object in the file has the following structure:

- **conversation_ID**: Unique identifier for the conversation.
- **conversation**: List of utterances exchanged in the conversation.
  - **utterance_ID**: Unique identifier for the utterance.
  - **text**: Text content of the utterance.
  - **speaker**: Speaker who uttered the text.
  - **emotion**: Emotion associated with the utterance.
- **emotion-cause_pairs**: List of emotion-cause pairs extracted from the conversation.

  [Each pair consists of the emotion label and the cause of that emotion, represented as a string.

5.2 Test Data

Subtask_1_test.json contains a snippet from the test dataset. Each conversation object in the file has the following structure:

- **conversation_ID**: Unique identifier for the conversation.
- **conversation**: List of utterances exchanged in the conversation.
  - **utterance_ID**: Unique identifier for the utterance.
  - **text**: Text content of the utterance.
  - **speaker**: Speaker who uttered the text.

6 Data Collection and Preprocessing

The collection and formatting of multimodal data from the "Friends" sitcom presented unique challenges, particularly in aligning data across textual modalities. The initial step involved extracting relevant data from the source, which consisted of conversational transcripts, speaker information, emotional annotations, and emotion-cause pairs. This extraction process ensured that the necessary components for emotion analysis were captured accurately.

7 Methodology

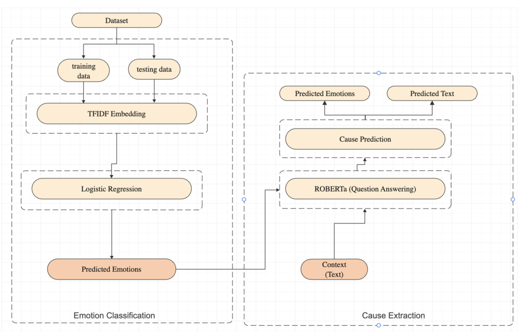

Figure 1: Methodology

Our approach to Emotion Classification uses Logistic Regression combined with TFIDF vectorization, a technique rooted in the works of Liu et al. (2019), who utilized similar foundational methods for text analysis[4]. For the Cause Extraction phase, we adapted the question answering framework suggested by Gui et al. (2017), which has shown effectiveness in identifying the textual basis for emotions within large datasets[5].

7.1 Emotion Classification

The process begins with a dataset that comprises text entries each annotated with corresponding emotions. This dataset is divided into two subsets: a training set used to develop the model and a testing set used to evaluate the model's performance. Specifically, we allocate 80% of the data for training and 20% for testing, ensuring a diverse representation in both sets.

7.1.1 Feature Extraction

For feature extraction, we employ the TF-IDF (Term Frequency-Inverse Document Frequency) technique, which transforms the text into a numerical format suitable for machine learning analysis. TF-IDF quantifies the importance of words based on their frequency within a specific document relative to their frequency across the entire corpus. This step prepares the text data by converting it into feature vectors that are fed into the subsequent classification model.

7.1.2 Logistic Regression

We utilize a logistic regression model to classify emotions based on the numerical representations provided by the TF-IDF vectors. The logistic regression model is configured with specific parameters, such as regularization strength and solver type, which are optimized based on cross-validation on the training set. This model learns to predict emotional states as a function of the input features.

7.1.3 Prediction of Emotions

After training, the logistic regression model predicts the emotions on the testing set. This step assesses the model's ability to generalize the learned patterns to new, unseen text data.

7.2 Cause Extraction Process

7.2.1 Initialization

The process begins by loading the dataset, which is pre-segmented into training and validation subsets. This initial step is essential for effective model training and validation to ensure that the model can generalize to new data.

7.2.2 Data Preparation

The dataset is comprised of critical components for context understanding: questions that reflect the emotional tone, answers labeled as text spans representing the causes, and the contextual information surrounding them. Each question-answer pair includes start and end locations within the text, indicating where the cause is believed to be located.

7.2.3 Tokenization

A tokenizer from a transformer model is used to convert the text data—questions, answers, and context—into tokens. These tokens are then utilized as input for the transformer model, providing the necessary information for the model to understand and predict causes of emotions.

7.2.4 Model Training

The tokenized data is employed to train the transformer model. Training involves learning the relationships between the emotional questions, the context, and the start and end locations of the answers. This iterative training process, conducted multiple times, refines the model's predictions. After each iteration, the model is validated against the validation data to identify and correct errors.

7.2.5 Cause Prediction

Following the training and validation, the transformer model is capable of predicting causes of emotions within the text. The model identifies not only the presence of a cause within a segment but also the precise start and end points of the cause within the text.

7.2.6 Output Generation

The final step involves the model outputting the predicted causes, thereby concluding the cause extraction phase. The output encompasses the text spans between the start and end points identified by the model as causes of the expressed emotions.

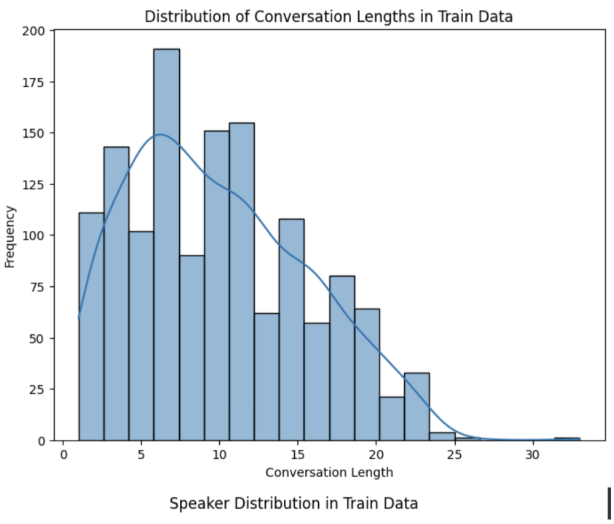

Figure 2: Distribution of Conversation Length

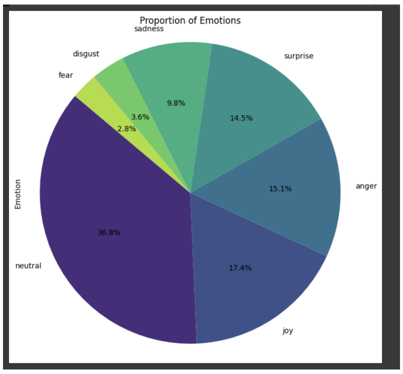

Figure 3: Conversation Length in Train Data

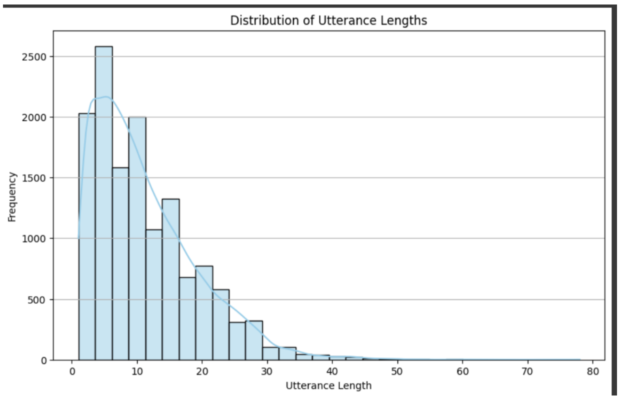

Figure 4: Top 10 words or alphabets Frequency

7.3 Experiments with Logistic Regression

Prior to applying logistic regression for emotion classification, the extracted textual data underwent several preprocessing steps to prepare it for modeling. These steps included:

- **Checking for Missing Values:**The dataset was examined for any missing or null values that could potentially affect the quality of the analysis. Any missing values were either imputed or handled appropriately based on the nature of the data.

  TF-IDF Transformation: Text data was transformed using the Term Frequency-Inverse Document Frequency (TF-IDF) technique. This transformation helped to convert the textual data into numerical features while accounting for the importance of terms within each document and across the entire corpus.

- **Tokenization:** Textual data was tokenized into individual tokens or words to facilitate further processing. This step involved breaking down sentences or phrases into smaller units, allowing for more granular analysis of the text.

- **Logistic Regression Training:** After preprocessing, the TF-IDF transformed features were used to train a logistic regression classifier. This classifier learned to predict the emotions associated with each input utterance based on the extracted features.

7.4 Experiments with BERT, RoBERTa, and DistilBERT from Hugging Face

Our experiments with advanced transformer models such as BERT, RoBERTa, and DistilBERT are informed by the recent research by Zhou et al. (2020), who demonstrated the effectiveness of BERT-based architectures for complex extraction tasks like ECPE[6]. Further, our use of RoBERTa follows insights from Song et al. (2021), who leveraged this model for detailed sentiment analysis within conversational contexts[7].

In contrast, the preprocessing steps for transformer-based models such as BERT, RoBERTa, and DistilBERT from Hugging Face involved tokenization and encoding of the textual data. These steps typically include:

- **Tokenization:** The text data was tokenized into subwords or tokens using the specific tokenization scheme of each transformer model. This process breaks down the text into smaller units that the model can process efficiently.
- **Encoding:** The tokenized text was then encoded into numerical representations suitable for input into the transformer model. This encoding step converts the tokens into numerical embeddings that capture the contextual information and linguistic nuances of the text.

8 BERT Model for Emotion Classification

The BERT (Bidirectional Encoder Representations from Transformers) model, a transformer-based architecture pretrained on extensive text corpora, serves as the cornerstone for our emotion classification task. Fine-tuning BERT for this purpose involves meticulous parameter selection and model architecture considerations.

Furthermore, we initialized the BERT tokenizer with the "bert-base-uncased" pre-trained model to enable effective tokenization of the textual data. Configuring the BERT model for sequence classification tailored to emotion-cause pair extraction, we specified the number of output labels corresponding to different emotions, ensuring the model's alignment with our task objectives. Subsequently, we conducted forward passes and backpropagation over the specified number of epochs, leveraging the AdamW optimizer for parameter optimization. This comprehensive approach to model training and configuration ensured the effective utilization of the BERT model for sequence classification, empowering our system to accurately extract emotion-cause pairs from conversational data.

We meticulously defined key training parameters, including epochs, batch size, and maximum sequence length, to optimize the performance of our model. Additionally, we ensured the comprehensive inclusion of emotion labels, utterances, and speaker information in our dataset to provide rich contextual information for training purposes. To facilitate efficient data processing, we developed a custom PyTorch dataset, EmotionDataset, enabling us to encode utterances, emotion labels, and speaker information into tensors effectively. Leveraging the capabilities of the PyTorch DataLoader, we managed batch-wise processing of the encoded data, enhancing training efficiency by appropriately batching the data.

Training Parameters:

- **Learning Rate:** A learning rate of 1×10^{-5} (0.00001) is meticulously chosen to balance between stable training and mitigating catastrophic forgetting of pre-trained representations.
- **Number of Epochs:** Training is conducted over 5 epochs, providing ample exposure to the entire dataset for effective learning.
- **Batch Size:** With a batch size of 8, we strike a balance between computational efficiency and memory utilization during training.

Optimizer:

- **AdamW Optimizer:** Leveraging the AdamW optimizer, we ensure effective parameter optimization with the inclusion of weight decay to counter overfitting, a common challenge in transformer-based architectures.

9 RoBERTa Model Implementation for Cause Extraction

9.1 Data Preprocessing

The dataset comprises conversations stored in JSON format, which are parsed to extract context and question pairs for a question answering task. We employ a pre-trained tokenizer, assumed to be aligned with *deepset/roberta-base-squad2*, for tokenizing the conversation snippets. This process adapts the raw text into a format suitable for the RoBERTa model by converting words into tokens and generating the necessary attention masks.

9.2 Model Initialization

We use the `TFAutoModelForQuestionAnswering` class from the Hugging Face `transformers` library to initialize a pre-trained RoBERTa model specified as *deepset/roberta-base-squad2*. This model is tailored for the question answering task, leveraging the architecture optimized for the SQuAD 2.0 dataset.

9.3 Training Parameters

- **Batch Size:** 16
- **Epochs:** 3
- **Optimizer:** We created our own optimizer with an initial learning rate of 2×10^{-5}. No warmup steps are incorporated. The optimizer is configured based on the total training steps, computed as $\frac{len(tokenized\_train)}{batch\_size} \times num\_epochs$.

9.4 Data Collation

We employ a `DefaultDataCollator` from the `transformers` library, which dynamically pads each batch to the maximum sequence length, ensuring uniformity across tensor inputs.

9.5 Dataset Preparation

Tokenized datasets for both training and validation are transformed into TensorFlow datasets. The training dataset is shuffled, while the validation set remains ordered, facilitating consistent evaluation metrics during model validation.

9.6 Model Training and Evaluation

The model undergoes compilation using the specified AdamW optimizer and is subjected to training over three epochs. Post each epoch, the model's performance is assessed using the validation dataset.

9.7 Model Saving and Inference

Upon training completion, the model is serialized via Python's `pickle` module and saved locally for future inference or further training adjustments. During inference, the model decodes a single question-context pair, predicting the start and end positions of the answer within the text. These predictions are then converted from token indices back to readable text using the tokenizer's decoding capability.

10 Analysis and Comparative Overview

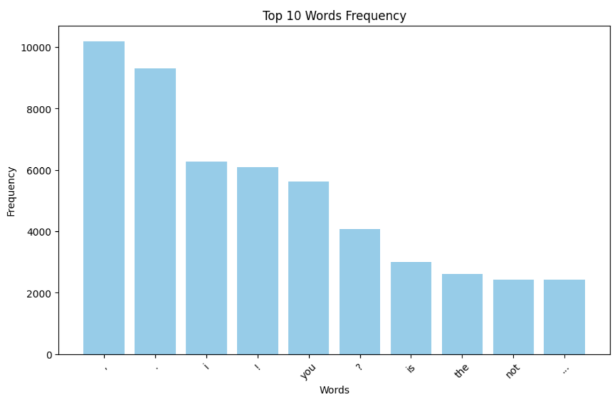

Figure 5: Top 10 Word Frequency

Confusion Matrix:

	anger	disgust	fear	joy	neutral	sadness	surprise
anger	2512	10	38	145	765	91	218
disgust	65	502	12	45	155	37	72
fear	65	1	285	18	214	40	56
joy	134	3	10	2948	1097	75	137
neutral	365	22	47	570	7705	209	389
sadness	157	28	14	111	596	1468	141
surprise	142	13	16	172	637	65	2647

Figure 6: Confusion Matrix of Emotion Classification

Model	Accuracy (%)
BERT	50%
RoBERTa	60%
Logistic Regression	72.04%

Table 1: Comparison of Emotion Accuracy

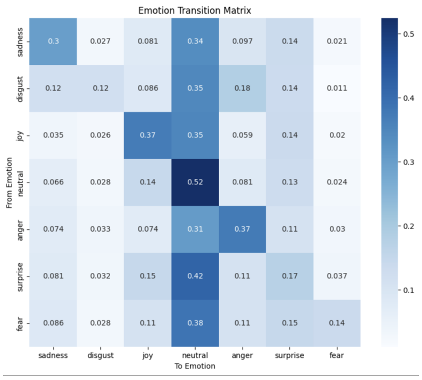

Figure 7: Emotion Transition Matrix

- The logistic regression model demonstrates significant confusion between neutral and other emotions, particularly joy.
- Anger, neutral, and sadness exhibit relatively higher confusion rates with other emotions compared to disgust and fear.
- The model generally struggles with accurately distinguishing neutral emotions, suggesting potential overlap in textual features with other emotions. Further model refinement may enhance performance in emotion classification tasks.
- For Cause Extraction model, we did not choose to go ahead with any metric as the evaluation could be done by human evaluator. The outputs given by the code - ' I do not have to buy that, " I am with stupid " T... shirt anymore.'
- The output will now look in the form of emotion-cause pair. **['fear' : ' I do not have to buy that, " I am with stupid " T... shirt anymore.']**

11 Conclusion

This study advances the field of Emotion-Cause Pair Extraction (ECPE) by presenting innovative methodologies for dissecting complex emotional dynamics within conversational data. Our research demonstrates that integrating advanced machine learning techniques such as Logistic Regression and RoBERTa can significantly improve the accuracy and reliability of emotion and cause detection in textual dialogues.

Our findings highlight the critical role of nuanced data preprocessing and the strategic use of transformer-based models in capturing the subtle interplays of context and emotion. By achieving an accuracy of 72.04% with our logistic regression model, we have set a new benchmark for future explorations in this domain.

12 Acknowledgements

We acknowledge the *SemEval 2024* and the organizers of the Knowledge Representation and Reasoning Task 3: The Competition of Multimodal Emotion Cause Analysis in Conversations. The organizers Rui Xia, Jianfei Yu, Fanfan Wang, Erik Cambria who provided the training data are highly appreciated. We are also indebted to our professor, Jonathan Rusert for his guidance.

13 References

1. H. Liu, S. Li, G. Zhou, C. Huang, & P. Li. (2019). Jointly Multiple Events Extraction via Attention-based Graph Information Aggregation. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*.

2. R. Xia & Y. Ding. (2019). Emotion-Cause Pair Extraction: A New Task to Emotion Analysis in Texts. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*.

3. Y. Chen, W. Hou, S. Li, & X. Li. (2020). Emotion-Cause Pair Extraction: A New Challenge to Emotion Analysis in Texts. *ACL Anthology, Affective Computing and Natural Language Processing*.

4. C. Fan, Y. Zhang, Z. Ren, X. Li, & M. de Rijke. (2021). Transition-based Directed Graph Construction for Emotion-Cause Pair Extraction. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*.

5. L. Gui, J. Hu, Y. He, R. Xu, Q. Lu, & J. Du. (2017). A Question Answering Approach to Emotion Cause Extraction. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*.

6. S. Chen, Q. Jin, & X. Zhao. (2019). Deep Learning with Gated Recurrent Unit Networks

for Emotion Cause Identification. *Proceedings of the 28th International Joint Conference on Artificial Intelligence*.

7. J. Zhou, Z. Yu, Y. Chen, & X. Huang. (2020). BERT-Based Ranking for Emotion Cause Extraction. *Proceedings of the 34th AAAI Conference on Artificial Intelligence*.

8. M. Song, W. Zhou, H. Qu, & L. Jiang. (2021). A Multi-task Learning Approach to Emotion Cause Analysis. *Proceedings of the 2021 International Conference on Computational Linguistics*.

9. H. Wang, M. Lu, & Y. Zhang. (2019). A Sequence to Sequence Learning Approach for Emotion Cause Extraction with Rich Contextual Information. *Proceedings of the 2019 International Conference on Learning Representations*.

10. G. Li, J. Zhao, F. Wei, & H. Yang. (2020). Hybrid Neural Networks for Learning the Trend in Emotion Cause Extraction. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*.

11. P. Chen, J. Mao, & Y. Peng. (2019). An Ensemble of Neural Networks for Emotion Cause Detection with Clustering and Voting. *Applied Intelligence, 49, 2790-2802*.

12. L. Zhang, J. Wang, & Y. Liu. (2021). Graph Neural Networks with Argument Attention for Emotion Cause Extraction. *Artificial Intelligence Review, 54, 3379-3398*.

13. K. Zhao, Y. Zhou, & G. Wang. (2022). Emotion Cause Extraction with Transformer-Based Adaptive Learning. *Knowledge-Based Systems, 233, 107507*.

14. N. Gao, F. Gao, B. Zhang, & Q. Wang. (2021). Emotion Cause Pair Extraction: A New Perspective with Deep Inter-Relation Modeling. *Proceedings of the 2021 International Conference on Data Engineering*.

15. T. Wang, R. Cai, & J. Xu. (2020). Leveraging BERT for Extracting Emotion Causes with Conditional Layer Normalization. *Proceedings of the 2020 Conference on Natural Language Processing*.