Emotion Cause Analysis in Conversations Using Textual Data

Mahisha Ramesh
MT23121
mahisha23121@iiitd.ac.in

Shaina Mehta
MT23139
shaina23139@iiitd.ac.in

Vatsal Lakhmani
2020148
vatsal20148@iiitd.ac.in

Mayank Jha
2020521
mayank20521@iiitd.ac.in

1. Introduction

In casual discussions, understanding human emotions and their underlying causes has drawn more attention in recent years. Accurately extracting emotion-cause pairs from text significantly impacts several disciplines, including affective computing, psychology, social robots, and human-computer interaction. Several causes lead to the motivation to investigate emotion-cause pair extraction in conversations:

1) **Enhanced Interaction between Humans and Computers**: As technology becomes more integrated into our daily lives, there is a rising need for more sympathetic and natural human-computer interactions. Consider a chatbot that helps consumers troubleshoot technical problems. The chatbot can identify signals of dissatisfaction or perplexity through text input analysis and provide customized solutions or further help, enhancing the user's overall experience and interaction efficacy.
2) **Developments in Affective Computing**: Affective computing aims to build systems that can identify, comprehend, and react to human emotions. Think of a text-based therapy platform where people go to get advice and emotional support. The platform can identify patterns suggestive of distress or anxiety by examining the emotional content of users' communications, including the language used and the subjects covered. It can then offer relevant resources or treatments to assist users in managing their emotions.
3) **Psychological insight**: Human behavior and interactions are greatly influenced by emotions. Conversational emotion-cause analysis can advance social science and psychology by illuminating the psychological processes that underlie human emotion regulation. Consider, In an online support group for individuals coping with mental health challenges, analyzing the text-based conversations can offer valuable insights into members' emotional states and the factors influencing their well-being. By extracting

emotion-cause pairs from the messages exchanged within the group, moderators and participants can better understand the underlying emotions and triggers, facilitating mutual support and constructive dialogue.

4) **Real-World Applications**: There are many real-world uses for the capacity to identify emotion-cause pairs in discussions automatically. Businesses can enhance service quality and customer happiness by comprehending the motivations and feelings of customers during customer service encounters. Analyzing the emotional content of discussions can also help therapists provide more individualized and successful solutions in mental health counseling.

1.1 Objective of the Project

This project aims to create a deep-learning model for emotion recognition in conversation (ERC), identifying emotion from the cause utterance in conversation and Cause pair identification in the conversations of textual data. This method will help different researchers create better AI-based conversational systems in the future.

2. Related Work

Several research studies have been done on emotional cause analysis. Earlier works are based on statistical-based learning methods and rule-based methods, mainly on emotion cause extraction, which has several limitations that are out of the report's scope. Nowadays, Deep Learning techniques are used for Emotion Cause Pair Extraction (ECPE), Emotion Recognition in Conversation (ERC) and emotion cause analysis related tasks. Xia et al. [1] proposed a hierarchical attention-based BiLSTM model for ECPE and ERC tasks. They trained the model on the dataset proposed by Gui et al. [2] by modifying it for their proposed tasks and achieved the F1 score of 61.28 percent on ECPE task and 83 percent on ERC task. Wang et al. [3] proposed the multimodal dataset for the ECPE task by modifying the MELD dataset proposed by [4] and [5] for the Emotion Recognition in Conversation task. They used their dataset to train the BiLSTm and CNN-based networks and achieved an F1 score of 51.32 percent. Wei et al. [6] proposed the one-step neural approach RANKCP for solving the ECPE problem. This model focused on modeling from a ranking perspective. They trained their model on the dataset proposed by [1] and achieved the F1 score of 66.10 percent, beating the baseline models. Fan et al. [7] proposed the transition-based framework to extract emotion based on directed graph construction. They trained their proposed models on the dataset proposed by [1] and achieved the F1 score of 67.99 percent. Mathur et al. [8] considered the ECPE task as the sequence labeling problem and proposed the models using different encoders, BiLSTM for adding contextual information of the conversation, and added CRF layer in the end to learn the inter-dependencies between adjacent utterances more effectively. They trained their models on a dataset proposed by [2] and [9] and achieved an F1 score of 17.59 percent on ECPE task, 54 percent on ERC task and

67 percent in cause pair identification task. Kazakov et al. [10] fine-tuned GPT-3.5 for emotion classification and performed the ECPE task using a BiLSTM-based neural network on the dataset proposed by [2] and [9] and achieved the weighted-average proportional F1 26.4 percent on ECPC task and 57 percent on ERC task. Wang et al. [11] proposed a Knowledge-Enhanced Hierarchical Transformers framework for the ECPE task and ERC task. They trained their model on the dataset proposed by [1] and achieved 77.79 percent on the ECPE task and 94.55 percent on the ERC task. Cheng et al. [12] proposed the Multimodal Emotion Recognition and Multimodal Emotion Cause Extraction (MER-MCE) framework using specialized emotion encoders such as Instruct ERC, HUBERT, expMAE, and Llama 2 model for ECPE task. They trained their model on the dataset proposed by [2] and [9] and achieved the weighted F1 score of 34.35 percent on the ECPE task and 68.07 percent on the ERC task. Kumar et al. proposed the Memory based Network and Transformer Based Architecture for ERC and Emotion Flip Reasoning tasks. They trained their models on the modified version of the dataset proposed by [4] and [5] and achieved better performance as compared to other models at that time.

3. Dataset

The dataset used in this project is from the competition SemEval 2024 Task 3 Subtask 1, proposed by [2] and [8]. This dataset is based on the American TV show Friends. The trial and testing part of the dataset does have labels. Hence, we will only use the training part of the dataset. As a result, the dataset consists of 1374 conversations, out of which we are taking 1099 conversations for training, 137 for validation, and remaining for testing.

4. Methodology

4.1 Data Pre-processing and Model Training of Emotion Recognition from the Cause Utterance

The following steps are followed for data pre-processing and training of Emotion Recognition from the Cause Utterance.

1. Emotion and cause pairs are extracted from the dataset and emotions are considered as labels and cause is considered as input. Then SMOTE technique is applied to prevent class imbalance.
2. Then the text is converted to its distributive representation using GloVe embeddings.
3. Then, the model is trained on Bi-LSTM with Attention with batch size of 32, using adam optimizer, cross entropy loss function till 20 epochs and its accuracy and F1 score is calculated.

4.2 Data Pre-processing and Model Training of Emotion Recognition in Conversation

The following procedure is the complete reimplementation of the paper [13]. The following steps have been followed for Emotion Recognition in Conversation (ERC) is as follows:

1. For each conversation, each user has been concatenated with the utterance. Here the ERC task is considered as the sequence labeling task.
2. Then the previous utterances and future utterances of the words have been taken and are concatenated with each other. These utterances are tokenized by the RoBERTa tokenizer in Huggingface.
3. Then they are trained on the RoBERTa model using AdamW optimizer, with batch size 40, cross entropy loss function, and learning rate of 0.01 till 10 epochs and its accuracy and F1 score is calculated.

4.3 Data Pre-processing and Model Training of Candidiate Cause Identification in Conversation

The following procedure is the complete reimplementation of the paper [13]. The following steps have been followed for Emotion Recognition in Conversation (ERC) is as follows:

- Utilized GloVe word embeddings for semantic representation.
- Preprocessed text data and labeled candidate cause or not based on emotion-cause pairs.
- Split the dataset into 70% training, 15% validation, and 15% testing.
- Constructed a feedforward neural network with two hidden layers of 64 neurons, ReLU activation, and dropout.
- Trained the model with binary cross-entropy loss and Adam optimizer over 10 epochs.
- Evaluated performance on the testing set for accuracy and predictions.

5. Results

All the models have been trained on Kaggle Notebook and Jupter Notebook. And their Accuracy and F1 Scores have been calculated. The classification report and the accuracy and F1 scores are given in table 1:

Task	Accuracy (%)	F1 Score (%)
Emotion Identification in Cause Utterance	44	44
Emotion Recognition in Conversation	58.03389830508474	43.087254031226385
Candidate Cause Identification	67	66

Table 1: Accuracy and F1 Scores of all the Tasks

```
Before SMOTE:
surprise: 2185 samples
anger: 2130 samples
sadness: 1443 samples
joy: 2760 samples
disgust: 534 samples
fear: 312 samples

After SMOTE:
surprise: 2760 samples
anger: 2760 samples
sadness: 2760 samples
joy: 2760 samples
disgust: 2760 samples
fear: 2760 samples
```
,

Figure 1:- Class imbalance removed using SMOTE for all emotions

```
104/104 ━━━━━━━━━━━━━━━━━━━━ 2s 22ms/step - accuracy: 0.4324 - loss: 1.9741
Test Accuracy: 0.4396135210990906
104/104 ━━━━━━━━━━━━━━━━━━━━ 3s 24ms/step
Classification Report on Test Data:
              precision    recall  f1-score   support

           0       0.59      0.62      0.61       582
           1       0.43      0.42      0.42       584
           2       0.46      0.43      0.44       554
           4       0.33      0.31      0.32       534
           5       0.36      0.42      0.39       533
           6       0.46      0.43      0.44       525

    accuracy                           0.44      3312
   macro avg       0.44      0.44      0.44      3312
weighted avg       0.44      0.44      0.44      3312
```

Figure 2:- Classification report of Emotion Identification

```
Classification Report on Test Data:
Cause: 7 i wish i did not have to move
True Emotion: surprise  Predicted Emotion: sadness

Cause: 11 yeah
True Emotion: surprise  Predicted Emotion: surprise

Cause: 6 i won it fair and square
True Emotion: anger     Predicted Emotion: sadness

Cause: 13 why do not we all go get something to eat
True Emotion: joy       Predicted Emotion: anger

Cause: gonna picked know 6 good
True Emotion: sadness   Predicted Emotion: disgust

Cause: 2 nice sidestep on the do do thing
True Emotion: joy       Predicted Emotion: joy
```

Figure 3:- True emotion vs Predicted emotion

```
64/64 ━━━━━━━━━━━━━━━━━━━━ 0s 1ms/step - accuracy: 0.6768 -
Test Accuracy: 0.6691140532493591
64/64 ━━━━━━━━━━━━━━━━━━━━ 0s 2ms/step
              precision    recall  f1-score   support

           0       0.71      0.72      0.71      1169
           1       0.61      0.61      0.61       874

    accuracy                           0.67      2043
   macro avg       0.66      0.66      0.66      2043
weighted avg       0.67      0.67      0.67      2043
```

Figure 4:- Candidate cause or not classification

From the above tables, one can infer that the performance of all the models is not much upto the mark as well as it is not much suitable for the real-time applications.

6. Conclusion and Future Work

This project discusses emotion recognition in conversation, identifying emotion from the cause utterance in conversation and Cause pair identification in the conversations of textual data using deep learning techniques. From the results one can infer that the performance of all the models are not upto the mark and are not suitable for real time applications. In the future, we are going to improve upon it by using Large Language Models plus we are going to perform ECPE tasks on real time dataset as well as used these rules for building the conversational bot.

References:

[1] Rui Xia and Zixiang Ding. 2019. Emotion-Cause Pair Extraction: A New Task to Emotion Analysis in Texts. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 1003–1012, Florence, Italy. Association for Computational Linguistics.

[2] Lin Gui, Dongyin Wu, Ruifeng Xu, Qin Lu, and Yu Zhou. 2016. Event-Driven Emotion Cause Extraction with Corpus Construction. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 1639–1649, Austin, Texas. Association for Computational Linguistics.

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

[4] S. Zahiri and J. D. Choi. Emotion Detection on TV Show Transcripts with Sequence-based Convolutional Neural Networks. In The AAAI Workshop on Affective Content Analysis, AFFCON'18, 2018.

[5] S. Poria, D. Hazarika, N. Majumder, G. Naik, E. Cambria, R. Mihalcea. MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversation. ACL 2019.

[6] Penghui Wei, Jiahao Zhao, and Wenji Mao. 2020. Effective Inter-Clause Modeling for End-to-End Emotion-Cause Pair Extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3171–3181, Online. Association for Computational Linguistics.

[7] Chuang Fan, Chaofa Yuan, Jiachen Du, Lin Gui, Min Yang, and Ruifeng Xu. 2020. Transition-based Directed Graph Construction for Emotion-Cause Pair Extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3707–3717, Online. Association for Computational Linguistics.

[8] Vardhan Mathur, S., Jindal, A. R., Mittal, H., & Shrivastava, M. (2024). LastResort at SemEval-2024 Task 3: Exploring Multimodal Emotion Cause Pair Extraction as Sequence Labelling Task. *arXiv e-prints*, arXiv-2404.

[9] Wang, F., Ma, H., Xia, R., Yu, J., & Cambria, E. (2024, June). SemEval-2024 Task 3: Multimodal Emotion Cause Analysis in Conversations. Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024), 2022–2033. Retrieved from https://aclanthology.org/2024.semeval2024-1.273

[10] Kazakov, R., Petukhova, K., & Kochmar, E. (2024). PetKaz at SemEval-2024 Task 3: Advancing Emotion Classification with an LLM for Emotion-Cause Pair Extraction in Conversations. *arXiv preprint arXiv:2404.05502*.

[11] Wang, Y., Li, Y., Yu, K., & Hu, Y. (2023, May). Knowledge-Enhanced Hierarchical Transformers for Emotion-Cause Pair Extraction. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 112-123). Cham: Springer Nature Switzerland.

[12] Cheng, Z., Niu, F., Lin, Y., Cheng, Z. Q., Zhang, B., & Peng, X. (2024). MIPS at SemEval-2024 Task 3: Multimodal Emotion-Cause Pair Extraction in Conversations with Multimodal Language Models. *arXiv preprint arXiv:2404.00511*.

[13] Kim, T., & Vossen, P. Emoberta: Speaker-aware emotion recognition in conversation with roberta. arXiv 2021. *arXiv preprint arXiv:2108.12009*.

[14] Kumar, S., Shrimal, A., Akhtar, M. S., & Chakraborty, T. (2022). Discovering emotion and reasoning its flip in multi-party conversations using masked memory network and transformer. *Knowledge-Based Systems*, *240*, 108112.