École Polytechnique

BACHELOR THESIS IN COMPUTER SCIENCE

Emotion Recognition in Conversation through Emotion Flow

Author:

Bruno Iorio, École Polytechnique

Advisor:

Gaël Guibon, LIPN - Université Sorbonne Paris Nord

Academic year 2024/2025

Abstract

Emotion Recognition in Conversation (ERC) is a very important domain, which has been gaining more attention in recent years, especially within NLP. In the scope of Emotion Recognition, identifying emotions in dialogues plays an essential role. This is because most of the emotional text data collection happens in the context of a conversation between two or more parties (e.g. customer service survey). In this paper we discuss the limitation of previous approches to the ERC task, while also evaluating an original approach to the same problem using Causal learning, which we identify as the Emotion Flow and attention mechanisms. (initial version)

Contents

1 Introduction

Emotions can be defined as one's psicological state, which can be caused by internal or external factors of an individual. While studying internal factors for the emotions, such as mental state and background, can be very difficult, external factors tend to be much more viable for studying, particularly Within the field of Natural Language Processing (NLP). The textual context provides very significant information that can directly influence the emotion present in a text. In a dialogue, for example, emotions in dialogues will depend heavely on the way that the dialogue progress.

(INSERT DIAGRAM EXEMPLIFYING THE ABOVE)

⋮

⋮

Emotion Recognition in Conversation (ERC) is a growing field within NLP which tackles the task of identifying emotions in conversation. Its increasing importance is partially explained by its multiple applications in different sectors of society. For example, the need of companies to understand and evaluate customer's satisfaction, which allows them to offer better services, being more competitive in the market. Moreover, the collection of this kind of data through a conversation framework allows us to capture emotions that are highly dependent on textual context, speaker's profile, etc. This allows models to learn how subtleties in speach affect emotions.

Even from a human perspective, identifying emotions can be very difficult, especially when we cannot rely on other factors such as voice tone, gestures, facial expressions etc. Also, even the classification of emotions can be considered ambiguous sometimes. For instance, joyfulness, happiness, and love are generally considered similar, and sometimes only being distinguished in terms of intensity. This disambiguouity is what imposes the challenges in ERC, motivating many different approaches for this task.

Previous researches have extensively studied how different model architectures affect the performance in the ERC task. DialogueRNN [5] uses Gated Recurrent Networks (GRN) to infer a representation for each emotion, while applying an attention mechanism to predict an emotion. ERC-DP [11] uses a BERT encoder, and introduces a personality state vector that incorporates information about the speaker itself for the emotion prediction.

⋮

⋮

Among the factors that can influence emotion detection, we particularly study the Emotion Flow in the conversation, which can be denoted as the graph describing the evolution, or change, in emotion throughout the conversation. However, the use of the Emotion Flow is limited to an utterance level. Which, while being a very logical way of studying emotions in conversation, has left some space for how a word level approach would perform. This is, instead of predicting an emotion for each utterance, a word level approach would predict an emotion for word. Indeed, to the extend of our research, there isn't any work evaluating the former approach in the task of ERC, leaving us with some room to study it.

In this Project, we aim at evaluating how a word level approach can be combined with the Emotion Flow in the ERC task. This comes with the challenge of determining ways to fuse the information given by the Emotion Flow and the flow of words.

⋮

⋮

(ADD PICTURE DEPICTING EMOTION FLOW)

2 Related Work

Probably, the most challenging problem in this project is to find efficient ways to fuse the textual and emotional information. In [4], it is applied a window transformer - using narrow 2D window masks - to catch short-term inter-utterance relations, used in the task of Causal Emotion Entailment(CEE), e.g. determining which parts of the text are causing each emotion. This was particularly inspired from MPEG [1], where the emotional information is embeddeded and later on fused with the textual information using an attention layer. This is useful because the CEE is closely related to ERC , and actually its use can actually improve the effeciency of ERC models [4]. In

3 Datasets used

Througout this project, we considered the following options of datasets:

DailyDialog: [3] Dataset containing 14,118 dialogues representing daily life dialogue situations. Each utterance is classified with one of the following emotions: anger, disgust, fear, happiness, sadness, surprise or other(no emotion).

MELD: [8] Dataset containing emotion anotation for over 1400 dialogues, and 13000 utterances, which were extract from the Friends TV show. It classifies each utterance as one of the following emotions: anger, disgust, fear, joy, neutral, sadness or surprise.

EmoryNLP: [13] 12,606 utterances extracted from the Friends TV show. It classifies each utterance as one of the following emotions: joyful, peaceful, powerful, scared, mad or sad.

After careful consideration, we decided to mainly focus on the EmoryNLP dataset. The reason for this is that it is considerably more balanced than the other two datasets. As we can see in the Diagrams below:

Insert Diagram for DD, MELD, EmoryNLP

4 Our Approach / Methodology (Absolutely going to change this title)

5 Results

6 Limitation of our approach

7 Conclusion

8 Perspectives

9 References

[1] Tiantian Chen, Ying Shen, Xuri Chen, Lin Zhang, and Shengjie Zhao. Mpeg: A multi-perspective enhanced graph attention network for causal emotion entailment in conversations. *IEEE Transactions on Affective Computing*, 15(3):1004–1017, 2024.

[2] Jiang Li, Xiaoping Wang, and Zhigang Zeng. A dual-stream recurrence-attention network with global–local awareness for emotion recognition in textual dialog. *Engineering Applications of Artificial Intelligence*, 128:107530, February 2024.

[3] Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. Dailydialog: A manually labelled multi-turn dialogue dataset, 2017.

[4] Hao Liu, Runguo Wei, Geng Tu, Jiali Lin, Dazhi Jiang, and Erik Cambria. Knowing what and why: Causal emotion entailment for emotion recognition in conversations. *Expert Systems with Applications*, 274:126924, 2025.

[5] Navonil Majumder, Soujanya Poria, Devamanyu Hazarika, Rada Mihalcea, Alexander Gelbukh, and Erik Cambria. Dialoguernn: An attentive rnn for emotion detection in conversations, 2019.

[6] Tao Meng, Fuchen Zhang, Yuntao Shou, Hongen Shao, Wei Ai, and Keqin Li. Masked graph learning with recurrent alignment for multimodal emotion recognition in conversation, 2024.

[7] Patrícia Pereira, Helena Moniz, and Joao Paulo Carvalho. Deep emotion recognition in textual conversations: A survey, 2024.

[8] Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. Meld: A multimodal multi-party dataset for emotion recognition in conversations, 2019.

[9] Xiangyu Qin, Zhiyu Wu, Jinshi Cui, Tingting Zhang, Yanran Li, Jian Luan, Bin Wang, and Li Wang. Bert-erc: Fine-tuning bert is enough for emotion recognition in conversation, 2023.

[10] Armand Stricker and Patrick Paroubek. A unified approach to emotion detection and task-oriented dialogue modeling, 2024.

[11] Xiaohan Wang, Linchao Zhu, and Yi Yang. T2vlad: Global-local sequence alignment for text-video retrieval, 2021.

[12] Yan Wang, Bo Wang, Yachao Zhao, Dongming Zhao, Xiaojia Jin, Jijun Zhang, Ruifang He, and Yuexian Hou. Emotion recognition in conversation via dynamic personality. In Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue, editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 5711–5722, Torino, Italia, May 2024. ELRA and ICCL.

[13] Sayyed M. Zahiri and Jinho D. Choi. Emotion detection on tv show transcripts with sequence-based convolutional neural networks, 2017.

A Appendix