Leveraging BERT for Enhanced Sentiment Analysis and Emotion-Cause Pair Detection in Conversational Contexts

Sergiu Craioveanu
University of Bucharest
sergiu-ionut.craioveanu@s.unibuc.ro

Abstract

In the rapidly evolving field of Natural Language Processing (NLP), understanding the nuances of human emotion in text, particularly in dialogues, poses a significant challenge. This paper introduces a novel approach to two critical subtasks in NLP: multiclass sentiment analysis and emotion-cause pair detection. We first develop an advanced sentiment analysis model capable of discerning multiple emotional states from textual data. Building on this foundation, we then address the more complex task of emotion-cause pair detection in dialogues. For this, we leverage the capabilities of BERT (Bidirectional Encoder Representations from Transformers) in Next Sentence Prediction (NSP), adapting it to identify causal relationships between emotional expressions and their triggers within conversations. Our methodology is applied to the Friends Conversation dataset, a rich source of colloquial and emotional dialogues. Through this application, we demonstrate how BERT's NSP feature, primarily used for predicting sentence sequences, can be innovatively adapted for detecting emotion-cause pairs in dialogue.

1 Introduction

Recent advancements in Natural Language Processing (NLP) have opened new frontiers in understanding complex language structures, particularly in sentiment analysis and emotion-cause recognition in conversational contexts. This paper presents a novel approach to these two intertwined subtasks, focusing on multiclass sentiment analysis and emotion-cause pair detection in dialogue systems. Our study leverages the capabilities of Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2018), specifically its Next Sentence Prediction (NSP) feature, to enhance the understanding of conversational dynamics in the Friends Conversation dataset, a rich source of colloquial and emotional dialogue.

Initially, we developed a multiclass sentiment analysis model that classifies dialogue into various emotional categories, surpassing previous benchmarks in accuracy and reliability. This model serves as the foundational layer for further analysis of emotional causality in conversations. The primary contribution of this work, however, lies in the novel application of BERT's NSP for detecting emotion-cause pairs in dialogues. By treating each conversational exchange as a 'sentence pair', our methodology utilizes NSP to infer the likelihood of causal relationships between non-consecutive utterances, thereby identifying the cause of specific emotional responses.

Our experiments demonstrate the efficacy of this approach, yielding significant improvements over traditional methods in emotion-cause pair detection. The model shows a remarkable ability to discern subtle emotional nuances and their triggers in natural conversation, a task that has traditionally posed challenges in computational linguistics.

The implications of this research are manifold. Not only do we build upon existing powerful sentiment analysis models, we bring a novel approach towards causal emotion-cause pair detection and fine-tuning BERT, which is relatively easy to train even on mainstream infrastructure (such as RTX 3090, in our case).

2 Related Work

Emotion cause extraction (ECE) was originally proposed by (Yat et al.). In their seminal study, (Chen et al., 2010), building upon the corpus introduced by (Yat et al.), posited that the granularity of a clause might be more conducive for annotating causality in the context of emotions. This proposition was a departure from traditional methods, advocating for the extraction of emotional causes at the clause level. However, they underscored two principal deficiencies in the Emotion-Cause Extraction (ECE) paradigm. Firstly, the prerequisite of manual emotion annotation prior to the extraction of causes presents a significant impediment to its pragmatic deployment, impinging upon the task's scalability and utility. Secondly, the sequential methodology of annotating emotions followed by cause extraction fails to acknowledge the inherent bidirectional relationship between emotions and their causative factors.

To rectify these methodological constraints, (Xia and Ding) put forth an innovative task delineated as Emotion-Cause Pair Extraction (ECPE). This avant-garde approach aims to simultaneously extract potential emotions and their corresponding causes from textual sources, thereby capturing the intricate interplay between these two entities. This conceptual advancement not only addresses the limitations of the ECE task but also enriches the field's understanding of the dynamic interrelations within emotional discourse. The introduction of the ECPE framework by (Xia and Ding) represents a pivotal development in this sphere of study, offering a more holistic and integrated perspective in the extraction of emotion-cause pairs from textual corpora.

3 Methodology

In this chapter, we explore the approaches and techniques employed to address the dual challenges of sentiment analysis and emotion-cause detection. Firstly, we present an in-depth examination of the dataset, including its origin, composition, and the preprocessing steps undertaken to render it suitable for our analysis. Following this, we shift our focus to the sentiment analysis task. Here, we detail the theoretical underpinnings of our approach, elucidating the rationale behind the selection of specific models and algorithms. Subsequently, we tackle the more complex task of emotion-cause detection. This segment of the chapter is dedicated to outlining our innovative approach to identifying and extracting emotion-cause pairs from textual data.

3.1 Sentiment Analysis

Given our Friends Dataset, we leverage a state-of-the-art text emotion classifier which has the backbone of DistilRoBERTa (Sanh et al.), whilst being pre-fine-tuned on the Emotion English Dataset (Hartmann, 2022).

Input generation relies on taking each utterance from each conversation, and passing it through the DistilRoBERTa tokenizer. This will yield attention masks and input_ids for each sentence. Afterwards, we will set the model to train mode and let it associate a sentence with the label belonging to the ground-truth emotions through further fine-tuning. The batch size is 128, optimizer is AdamW with learning rate 2e-5, and tokenizer has the default parameters.

In the below images, we can see the emotion distribution for each split. The train input contains 10246 samples, whereas the test input contains 3373 samples. Below are the emotion distributions:

Emotion	Count
Anger	1241
Disgust	303
Fear	267
Joy	1707
Neutral	4464
Sadness	871
Surprise	1393

Table 1: Train Samples Emotion Distribution

Emotion	Count
Anger	374
Disgust	111
Fear	106
Joy	594
Neutral	1465
Sadness	276
Surprise	447

Table 2: Test Samples Emotion Distribution

3.2 Emotion-Cause Detection using BERT's NSP

The core of our methodology revolves around the employment of BERT for Next Sentence Prediction (NSP) to evaluate relational dynamics between sentences in a conversation. We initiate this process by constructing all possible causal sentence pairs within each conversation. These pairs are then classified into two categories: pairs that are present in the annotated ground truth are labeled as '1', indicating a true causal relationship; and pairs not found in the ground truth are labeled as '0', signifying a lack of causal relation.

An important refinement in our approach is the exclusion of pairs where both sentences are identical. The rationale behind this decision is to delegate the responsibility of identifying emotional content within a single sentence to the sentiment analysis model. This model is configured to recognize and include such sentences as pairs only if they exhibit an emotion other than neutrality.

The performance of the BERT-based NSP model is then evaluated using the F1 score based on its performance across all non-identical sentence pairs (e.g., pairs 1-2, 1-3, 1-4, 2-3, etc.). We will understand more on this within the experiment results.

This evaluation is performed without considering the span of the detected emotions or causes. The accuracy assessment encompasses: 1. The detection accuracy of emotion-cause pairs as determined by the NSP model; 2. The contribution of the sentiment analysis model, which, in our framework, is confined to the identification of self-matched sentences that exhibit a specific emotion.

By integrating these two assessment phases, our methodology not only leverages the predictive power of BERT in understanding sentence relationships but also complements it with the nuanced detection of emotional content through sentiment analysis. This combined approach presents a comprehensive framework for the automated and nuanced understanding of emotional causality in conversational datasets. We will also present the combined F1 score of this approach.

We had 61826 emotion-cause pairs, 3255 as true labels, and 58571 as false, respectively. These did not include inverse causal pairs (i.e. 9-8). That poses a task open for further research.

Similar to the previous task, during training we employ the default tokenizer, with a maximum sequence length of 256 with maximum length padding and truncation. Batch size was of 64 (this was the maximum the hardware could allow for), and the optimizer was AdamW with a learning rate of 1e-5.

4 Experiments

In the experimental analysis of the **michelle-jieli/emotion_text_classifier** model on HuggingFace, a series of trials were conducted to assess the model's performance across different training epochs. As depicted in Table 3, the model demonstrated varying degrees of efficiency in train and test accuracies over different epoch settings. With a brief training period of 2 epochs, the model achieved a train accuracy of 0.753 and a test accuracy of 0.679. However, extending the training to 20 epochs resulted in a significant increase in train accuracy, reaching 0.955, albeit with a slight decrease in test accuracy to 0.646, sign of extreme overfitting.

Epochs	Train Accuracy	Test Accuracy
2	0.753	0.679
20	0.955	0.646
5	0.966	0.645

Table 3: Performance of the Emotion Text Classifier Model

This seemed to be one of HuggingFace's most popular sentiment analysis models - for this scale at least. Given that LLMs can solve this problem considerably better, our purpose was not to re-invent the wheel on a potentially solved task, but to find a powerful solution within the reach of our limited resources. We consider this model to be a powerful foundation for our next task, which could perhaps open the way for further research into sentiment analysis on limited hardware, and with a limited number of model parameters.

Given the HuggingFace implementation of BERT, it was quite easy to setup the data for Next Sentence Prediction. The first measurements were on how well BERT was able to associate causal emotion-cause pairs within our dataset (i.e. 1-2, 1-3, etc.). This can be seen in the confusion matrices below.

		Predicted labels	
		0	1
True labels	0	58026	545
	1	65	3190

F1 Score: 0.9905

Table 4: Confusion Matrix for Causal Train Emotion-Cause Pairs

		Predicted labels	
		0	1
True labels	0	18915	825
	1	769	250

F1 Score: 0.9242

Table 5: Confusion Matrix for Causal Test Emotion-Cause Pairs

The results are nothing short of impressive - it's pretty clear that BERT can clearly surpass other state-of-the-art approaches (Wang et al.) with minimal training. Whilst we approach the task of causal and self-matching pairs, we leave room for further experimentation and research in terms of inverse-causal pairs. It is, however, our assumption that the model will perform well on that particular task, as it has on the current one.

As mentioned in the methodology, we manage to combine the text sentiment analysis model for self-matching pairs, as well as BERT's NSP for causal emotion-cause pairs, and obtain the following impressive results:

		Predicted labels	
		0	1
True labels	0	54396	1438
	1	4051	1941

F1 Score: 0.90

Table 6: Confusion Matrix for Causal and Self-Matching Train Emotion-Cause Pairs

		Predicted labels	
		0	1
True labels	0	20611	1353
	1	1276	892

F1 Score: 0.8919

Table 7: Confusion Matrix for Causal and Self-Matching Test Emotion-Cause Pairs

We notice a slightly decreased F1 score after taking into consideration both types of emotion-cause pairs - our ablation studies show that even if our sentiment analysis model would have been perfect, our F1 score increase would have been in the ranges of 3-4%. As such, it would seem like perhaps the largest benefit of all could be to double down on the BERT NSP approach and perfect it. Of course, should we have included inverse-causal pairs, the scores may look quite different altogether.

5 Conclusion

In conclusion, the methodology presented in this paper marks a significant advancement in the field of Natural Language Processing, particularly in the nuanced detection of emotion and their causal triggers in textual dialogues. Our research introduces a simple yet groundbreaking approach to the intricate tasks of multiclass sentiment analysis and emotion-cause pair detection. By harnessing the predictive power of BERT's Next Sentence Prediction feature, we have not only enhanced the model's capability to discern multi-faceted emotional states but have also laid the groundwork for sophisticated emotion-cause pair detection.

The efficacy of our model is evidenced by its successful application to the Friends Conversation dataset, which signifies a promising leap forward in the performance of NLP tasks on mainstream hardware. This innovative use of BERT for emotion-cause pair detection in dialogues is poised to catalyze remarkable improvements in the accuracy and efficiency of similar models. Moreover, the flexibility of our approach opens avenues for future research, including the exploration of inverse causal relationships between emotions.

The implications of our work could set a precedent for future NLP applications that require a deep understanding of emotional dynamics in text. We anticipate that our contributions will not only enhance the user experience in conversational AI systems but also extend to other domains where emotional intelligence in language understanding is paramount.

References

Ying Chen, Sophia Yat Mei Lee, Shoushan Li, and Chu-Ren Huang. 2010. Emotion Cause Detection with Linguistic Constructions. pages 179–187.

Jacob Devlin, Ming Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1:4171–4186.

Jochen Hartmann. 2022. Emotion english distilroberta-base. https://huggingface.co/j-hartmann/emotion-english-distilroberta-base/.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter.

Fanfan Wang, Zixiang Ding, Rui Xia, Zhaoyu Li, and Jianfei Yu. Multimodal Emotion-Cause Pair Extraction in Conversations.

Rui Xia and Zixiang Ding. Emotion-Cause Pair Extraction: A New Task to Emotion Analysis in Texts. pages 1003–1012.

Sophia Yat, Mei Lee, Ying Chen, and Chu-Ren Huang. A Text-driven Rule-based System for Emotion Cause Detection. pages 45–53.