Hidetsune at SemEval-2024 Task 3: A Simple Textual Approach to Emotion Classification and Emotion Cause Analysis in Conversations Using Machine Learning and Next Sentence Prediction

Hidetsune Takahashi
Waseda University
takahashi78h@toki.waseda.jp

Abstract

In this system paper for SemEval-2024 Task 3 subtask 2, I present my simple textual approach to emotion classification and emotion cause analysis in conversations using machine learning and next sentence prediction. I train a SpaCy model for emotion classification and use next sentence prediction with BERT for emotion cause analysis. While speaker names and audio-visual clips are given in addition to text of the conversations, my approach uses textual data only to test my methodology to combine machine learning with next sentence prediction. This paper reveals both strengths and weaknesses of my trial, suggesting a direction of future studies to improve my introductory solution.

1 Introduction

SemEval 2024 Task 3 (Wang et al., 2024) calls for assigning an emotion to each utterance and extracting its emotion cause in conversations. Subtask 2, which I participate in, requires emotion classification and identification of emotion cause utterances with audio-visual clips available whereas subtask 1 requires identification of specific textual span as well without audio-visual clips.

I participate in subtask 2, for which a speaker name, text and an audio-visual clip are given for each utterance. Instead of not identifying specific cause span in the emotion cause utterance in this subtask, I set a limitation to use textual data only while audio-visual data are also available. Therefore, my methodology uses textual data of utterances only as input to classify emotions and identify cause utterance numbers as output. For this reason, training data from subtask 1, for which video names are not given, are used instead of data from the subtask I participate in.

While the task (Wang et al., 2024) prohibits use of additional annotation data, I overlooked the sentence stating the rule and mistakenly used additional data for my solution. I would like to show my appreciation for the task organizers and readers acknowledging and understanding my mistake of using additional data.

For data preparation, official training data for subtask 1 (CSV converted version) (Wang et al., 2023), training data (translated and CSV converted version) from SemEval2024 Task10 subtask 1 (ERC) by Kumar et al. (2023) and data by Nikam (n.d.) are used. They all are concatenated in that order and adjusted so that the resulting dataset has first 7001 neutral utterances (including 7000th counting from 0) and first 5001 utterances at maximum for each emotion other than neutral.

Then, SpaCy-v3 model (Kömeçoğlu, 2023) is trained using the adjusted training data for emotion classification. In addition to that, next sentence prediction (Cathrine, 2023) is used for identification of emotion cause utterances. In that step, I decided to simplify the methodology by hypothesising that the emotion cause utterance is the utterance itself or its previous utterance. With this simple assumption, my algorithm checks the relatedness of each utterance and its previous utterance using next sentence prediction (Cathrine, 2023), which returns true or false. Previous utterance is chosen as cause utterance if the two utterances are deemed related, and the utterance itself if not.

The result shows a limited performance of my introductory solution, but it also clarifies a direction to its improvement. Although my combined methodology has a large room for improvement, it does have a potential in its simplicity and limitation to use textual data only. This paper aims to share an experimental trial to test my combined methodology, guiding a direction to its future application and improvement.

My code is available on GitHub [1].

[1] https://github.com/Hidetsune/SemEval2024_Task3.git

Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024), pages 361–364

2 Background

The subtask I participate in (subtask 2) focuses on emotion classification and emotion cause analysis with text data and audio-visual clips. Subtask 1, on the other hand, does not allow participants to use audio-visual clips and requires extracting specific textual cause spans as well. My participation in subtask 2 sets a limitation to use textual data only, which means that it is substantially the same as subtask 1 except that I do not extract specific textual cause spans. Training data from subtask 1 are used instead of that from subtask 2 because they seem to be identical to each other except that they have no video names in the dataset. Therefore, my methodology for subtask 2 uses data from subtask 1 and additional data from other sources for training. Given the evaluation dataset with audio-visual clips available, my methodology, which is trained by textual data only, assigns an emotion category and its cause utterance as output using textual data of the evaluation dataset.

This task is technically a mixture of two topics, which are emotion classification and emotion cause analysis. As for emotion classification, many previous studies have been conducted especially on social media including Twitter and Facebook. For instance, a work by Gaind et al. (2019) classifies text on social media into six emotion categories with high accuracy. Another study by Brynielsson et al. (2014) investigates in people's emotions during crises using a support vector machine. In addition to its use for social media, its application to real conversations is also getting an attention. A study by Graterol et al. (2021), for example, applies emotion detection to social robotics, aiming to improve its ability to interpret feelings of humans from a viewpoint of NLP methods.

There are many previous studies for emotion cause analysis too. A study by Fan et al. (2019), for example, uses hierarchical neural network to get high accuracy. Another study by Ding et al. (2020) adopts a complicated approach, resulting in reliable accuracy.

On the other hand, this paper aims to test a simple approach to combine classical machine learning method with next sentence prediction with a certain assumption. My methodology has a strength in its simplicity, but the result shows a large room for improvement.

3 System overview

The main strategy of my system is a combination of classical machine learning method with next sentence prediction. Machine leaning is used for emotion classification and next sentence prediction is used for identification of emotion cause utterances. Audio-visual clips are available in this subtask, but only textual data of the utterances are used for my solution. A quick overview of my combined methodology is as follows.

1. **Training data preparation**: Official training data from subtask 1 (Wang et al., 2023) are converted from a json file into a pandas dataframe. Similarly, training data from SemEval2024 Task10 subtask 1 (ERC) (Kumar et al., 2023) are translated into English and converted into a pandas dataframe. The converted dataframes and data by Nikam (n.d.) are concatenated to compose an adjusted training data. The adjusted data have two columns, in which text and an emotion are stored respectively for each utterance.

2. **Emotion classification using machine learning**: Using the adjusted training data, SpaCy-v3 model (Kömeçoğlu, 2023) is trained and used for emotion classification. It assigns an emotion to each utterance from "neutral", "surprise", "anger", "sadness", "joy", "disgust" and "fear".

3. **Emotion cause utterance identification using next sentence prediction**: If an assigned emotion is not "neutral", next sentence prediction (Cathrine, 2023) identifies its emotion cause utterance. My methodology works under the simple assumption that emotion cause utterance is the utterance itself or its previous utterance.

In the first step, training data are prepared from multiple sources. The official training data from subtask 1 (Wang et al., 2023) are imported as a json file and converted into a pandas dataframe with text and an emotion for each utterance. Here, data from subtask 1 are used instead of that from subtask 2 because it is likely that the data are identical to each other except that video names are not given for data of subtask 1. Then, the resulting pandas dataframe, translated and converted version of training data from SemEval2024 Task10 (Kumar et al., 2023) and data by Nikam (n.d.) are imported via CSV file format. As for the two additional datasets, irrelevant columns are dropped so that they are composed of two columns, in which text and emotions are stored respectively. They are concatenated into one dataframe and adjusted to have 7001 utterances (including 7000th counting from 0) for neutral and 5001 at maximum for each one of the other emotions related to this task (surprise, anger, sadness, joy, disgust and fear).

After this process of data preparation, SpaCy-v3 model (Kömeçoğlu, 2023) is trained using the prepared training data. An unlabeled evaluation dataset is imported as a json file, and the trained model assigns an emotion to each utterance.

At the same step, next sentence prediction with BERT (Cathrine, 2023) assigns an utterance number of emotion cause to each utterance if its assigned emotion is not "neutral". As stated before, it is hypothesised that the emotion cause utterance is either the utterance itself or its previous utterance. Under this assumption, next sentence prediction (Cathrine, 2023) checks whether or not an utterance that it is looking at is related to its previous utterance. The previous utterance is chosen as its emotion cause utterance if these are deemed related, and the utterance itself is chosen if not. After all these processes, lists that include emotions with the utterance numbers and emotion cause utterance numbers (['2_sadness', '1'] for example) are added to the original evaluation data for submission.

My participation in this task using the combined methodology reveals its limitations of the simple approach to emotion cause analysis. Since my methodology trains SpaCy-v3 model with over 33000 utterances, it is more natural to assume that the simplistic application of next sentence prediction is the main reason for the limited accuracy. My algorithm takes only the utterance itself and its previous utterance into account as possible emotion cause utterances. This premise does not allow my solution to cover cases where one utterance has an influence beyond multiple utterances, limiting the ability to deal with the entire conversation from a macroscopic point of view. On the other hand, there is no doubt that the accuracy of emotion classification is also a reason for the limited ability of my trial. Emotion cause utterances are assigned to non-neutral emotion utterances only, meaning that the algorithm loses its accuracy for both emotion classification and emotion cause analysis at a time if the trained model mistakenly assigns "neutral" to non-neutral emotion utterances.

4 Experimental setup

For the emotion classification part, multiple datasets needed to be processed to make an adjusted training dataset.

First of all, the official training data for subtask 1 (Wang et al., 2023) are imported as a json file. In the dataset, a conversation ID is assigned to each conversation, and one conversation has multiple utterances. For each utterance, an utterance ID, text, its speaker name and an emotion are assigned. The json file is converted into a pandas dataframe to make the data easier to deal with. Conversation IDs, utterance IDs and speaker names are dropped from the dataframe so that it has only "text" and "emotion" as columns. Data from SemEval2024 Task10 (Kumar et al., 2023), which have Hindi-English code-mixed utterances, are translated into English and converted into a pandas dataframe similarly. Data by Nikam (n.d.) are also imported as a pandas dataframe, and irrelevant columns of the two additional datasets are dropped so that they have text utterances and emotions as columns only. The number of utterances for each different emotion category is as shown on Table1. After these processes, the three dataframes (official training data for subtask 1, data from SemEval2024 Task10 (Kumar et al., 2023) and data by Nikam (n.d.)) are concatenated into one dataframe in that order. Only first 7001 (including 7000th counting from 0) neutral emotion utterances and first 5001 utterances for each one of the other emotions are extracted, dropping all the utterances that exceed the limitation from the concatenated dataset to compose an adjusted training data.

After this data preparation step, SpaCy-v3 model (Kömeçoğlu, 2023) is trained using the adjusted training data, and the trained model is used for emotion classification of the unlabeled evaluation dataset. Next sentence prediction (Cathrine, 2023) is also used for emotion cause analysis as stated in the previous section.

5 Results

In the evaluation phase, my solution was tested using the unlabeled test dataset. The result shows a limited ability of my approach, which combines classical machine learning with next sentence prediction under a simplistic assumption. The scores

Dataset	Anger	Disgust	Fear	Joy	Neutral	Sadness	Shame	Surprise	Contempt
Data from subtask1	1615	414	373	2301	5929	1147	0	1840	0
Data from Task10	819	127	514	1596	3909	558	0	441	542
Data by Nikam	4286	856	5409	11037	1811	6719	146	4062	0

Table 1: Datasets and emotion categories

w-avg. F1	F1	Ranking
0.1288	0.1389	12/16

Table 2: Task scores in evaluation phase

are displayed in Table 2.

6 Conclusions

To summarize, my methodology sets a limitation to use textual data only, testing a simple algorithm with a certain premise. I use classical machine learning for emotion classification, and next sentence prediction for identification of emotion cause utterances.

The next sentence prediction (Cathrine, 2023) in my simple approach takes only the utterance itself and its previous utterance into account, limiting its ability to cover the entire conversation from a macroscopic viewpoint. In addition to that, the accuracy of emotion classification between neutral and non-neutral turned out to be more important than previously thought since it has a significant effect on identification of emotion cause utterances as well.

Although the trial of my simple approach has a large room for improvement, it clearly guides a direction to its future studies. With improvements to enhance the ability to cover conversations from a macroscopic point of view, it might open the door for the potential of my combined methodology.

References

Joel Brynielsson, Fredrik Johansson, Carl Jonsson, and Anders Westling. 2014. Emotion classification of social media posts for estimating people's reactions to communicated alert messages during crises. *Security Informatics*, 3:1–11.

Jeeva Cathrine. 2023. Next sentence prediction with BERT. *scaler*.

Zixiang Ding, Rui Xia, and Jianfei Yu. 2020. End-to-end emotion-cause pair extraction based on sliding window multi-label learning. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, pages 3574–3583.

Chuang Fan, Hongyu Yan, Jiachen Du, Lin Gui, Lidong Bing, Min Yang, Ruifeng Xu, and Ruibin Mao. 2019. A knowledge regularized hierarchical approach for emotion cause analysis. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5618–5628. Association for Computational Linguistics.

Bharat Gaind, Varun Syal, and Sneha Padgalwar. 2019. Emotion detection and analysis on social media. *arXiv preprint arXiv:1901.08458*.

Wilfredo Graterol, Jose Diaz-Amado, Yudith Cardinale, Irvin Dongo, Edmundo Lopes-Silva, and Cleia Santos-Libarino. 2021. Emotion Detection for Social Robots Based on NLP Transformers and an Emotion Ontology. *Sensors*, 21(4).

Shivani Kumar, Ramaneswaran S, Md Akhtar, and Tanmoy Chakraborty. 2023. From multilingual complexity to emotional clarity: Leveraging commonsense to unveil emotions in code-mixed dialogues. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 9638–9652, Singapore. Association for Computational Linguistics.

Başak Kömeçoğlu, Buluz. 2023. Emotion Classification with SpaCy v3 Comet.

Sanket Nikam. n.d. Emotion Detection in Text using Natural Language Processing.

Fanfan Wang, Zixiang Ding, Rui Xia, Zhaoyu Li, and Jianfei Yu. 2023. Multimodal emotion-cause pair extraction in conversations. *IEEE Transactions on Affective Computing*, 14(3):1832–1844.

Fanfan Wang, Heqing Ma, Rui Xia, Jianfei Yu, and Erik Cambria. 2024. Semeval-2024 task 3: Multimodal emotion cause analysis in conversations. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pages 2022–2033, Mexico City, Mexico. Association for Computational Linguistics.