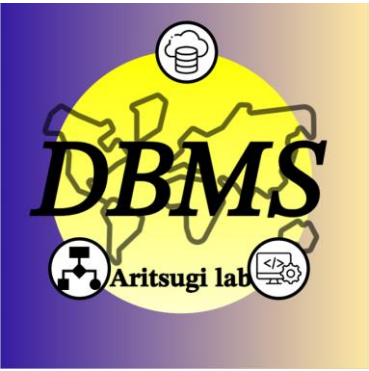


Prompt based zero shot emotion classification

Defense of Master thesis

Gita Tabe Jamaat

Student ID: 237-D8905

Supervisor: Professor Masayoshi Aritsugi

outline

- Introduction
- Background
- Research methods
- Experiments, results and analysis
- Conclusion and future work

Introduction

Introduction

Recent research works on emotion classification

- need predefined labels for classification.
- It’s not possible to determine classes that have not been seen during training

Solution: zero shot classification

- Checking the entailment of text and emotions using pre-trained model
- It’s possible to use descriptions or synonyms of emotions

Problem statement

The recent studies have disadvantages because of training on datasets with specific labels [1].

The idea of zero shot classification can solve this problem by considering emotion as the text and doing textual emotion classification

Therefore, we can add templates at the beginning of emotions which are called prompts or use synonym of emotions by different strategies

[1] Angeliki Lazaridou, Georgiana Dinu, and Marco Baroni. 2015. "Hubness and Pollution: Delving into Cross-Space Mapping for Zero-Shot Learning". In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 270 280, Beijing, China. Association for Computational Linguistics.

Background

Emotion classification

- Emotion classification is one of the significant works in natural language processing
- It maps textual units to emotion classes
- The emotion classes can be determined based on existing emotions in datasets

Zero shot learning

- Determining classes that have not been seen during training
- Therefore, it’s a kind of unsupervised method.
- We can obtain a model which doesn’t need any labeled training data.
- One approach for zero shot classification can be sequence pair classification

Natural language inference for zero shot emotion classification

- It is based on checking the entailment of 2 sentences
- Therefore, the emotion can be considered as a text
- The emotion can be represented in different descriptions or different terms
- The models can check the entailment of emotions and texts

Research methods

Proposed solution

- The previous studies have shown the importance of prompt to enhance zero shot emotion classification by using pre-trained models and using prompts of emotions and synonyms of emotions [2]
- These studies have limitation in showing the effects of prompts clearly and designing the guided prompts
- In this work, we have analyzed the results of using prompts and use more strategies of using synonyms of emotions

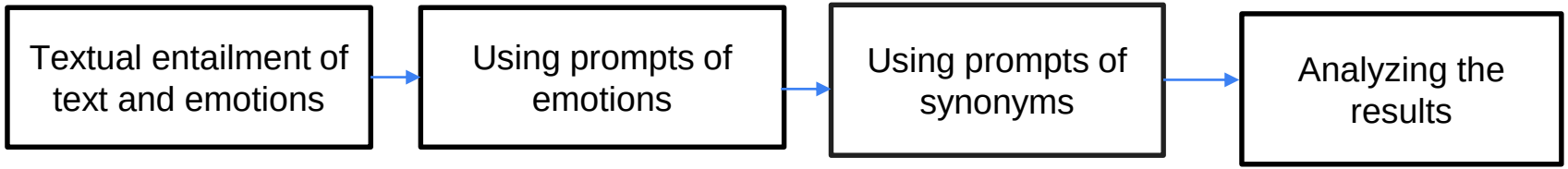


Figure 1: The framework of proposed method

[2] Flor Miriam Plaza-del-Arco, María-Teresa Martín-Valdivia, and Roman Klinger. 2022. Natural Language Inference Prompts for Zero-shot Emotion Classification in Text across Corpora. In Proceedings of the 29th International Conference on Computational Linguistics, pages 6805–6817, Gyeongju, Republic of Korea. International Committee on Computational Linguistic

Natural language task

- It is based on checking the entailment of 2 sentences

  function : $f_{(s1,s2)} \rightarrow \{C, N, E\}$

  s1: premise s2: hypothesis C: contradiction ,N: neutral output, E: entailment

Figure 2: Textual entailment

The entailment shows how much the text to be classified and emotions are related to each other.

- Only the scores of contradiction and entailment are considered as a vector
- The unnormalized scores of contradiction and entailment are converted to probabilities
- The emotion with the highest entailment probability is considered as the predicted emotion

Prompt generation

Since the emotion can be considered as text in zero-shot classification, it's possible to add templates at the beginning of the emotions and use different descriptions of emotions which are called prompts [3].

Table 1: prompts of emotions

ID	prompt	example
emotion name	emotion name	disgust
emotion prompt	"The emotion is"+ emotion	The emotion is disgust
expr emo	"This text expresses"+ emotion	This text expresses disgust
feels emo	"This person feels"+ emotion	This person feels disgusted

[3] Wenpeng Yin, Jamaal Hay, and Dan Roth. 2019. Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach. 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 3914–3923, Hong Kong, China. Association for Computational Linguistics.

Using synonyms in prompts of emotions

It’s also possible to use different synonyms of emotions.
So, we can replace the synonyms of emotions in prompts and add templates at the beginning of synonyms [4].

Table 2: prompts of synonyms

ID	Prompt	example
emo-s	Emotion synonym	happy
expr-s	“This text expresses”+emotion synonym	This text expresses happiness
feels-s	“This person feels”+emotion synonym	This person feels happy
emotion-s	“The emotion is”+ emotion synonym	The emotion is happiness

[4] Bareiss Patrick; Klinger Roman; Barnes Jeremy. 2024: English Prompts are Better for NLI-based Zero-Shot Emotion Classification than Target-Language Prompts, in: WWW ’24: Companion Proceedings of the ACM on Web Conference 2024, New York,

Synonyms of emotions

Table 3: synonyms of emotions

emotion	synonyms
sadness	sadness, unhappiness, grief, sorrow, loneliness, depression
joy	joy, an achievement, pleasure, the awesome, happiness, the blessing
anger	anger, annoyance, rage, outrage, fury, irritation
disgust	disgust, loathing, bitterness, ugliness, repugnance, revulsion
fear	fear, horror, anxiety, terror, dread, scare
surprise	surprise, astonishment, amazement, impression, perplexity, shock
shame	shame, humiliation, embarrassment, disgrace, dishonor, discredit
guilt	guilt, culpability, responsibility, blameworthy, misconduct, regret

Strategies to use synonyms of emotions

- Checking the entailment of different synonyms for each emotion and getting average of their entailment probabilities as the entailment probability of the emotion
- Checking the entailment of different synonyms for each emotion and setting maximum of them as the entailment probability of the emotion
- Combining the results of two methods by getting average of the entailment probabilities of each emotion in two results of two previous methods

Experiment, results and analysis

Datasets and Models

Datasets:

- Isear dataset: 7515 data Labels: anger, disgust, fear, guilt, joy, sadness, shame
- Tec dataset: 21051 data Labels: anger, disgust, fear, joy, sadness, surprise

Preprocessing datasets before doing classification:

Removing some signs such as ; , :, *, # , = , (,), & and integers

Pre_trained Bert based models (publicly available within hugging face python library):

- Deberta: microsoft/deberta-v2-xlarge-mnli
- Roberta: roberta-large-mnli,
- Bart: facebook/bart-large-mnli

Results of prompts of emotions (Isear dataset)

- When using Deberta model, emotion prompt and expr emo could lead to improvement
- When using Bart model, expr emo could lead to improvement.

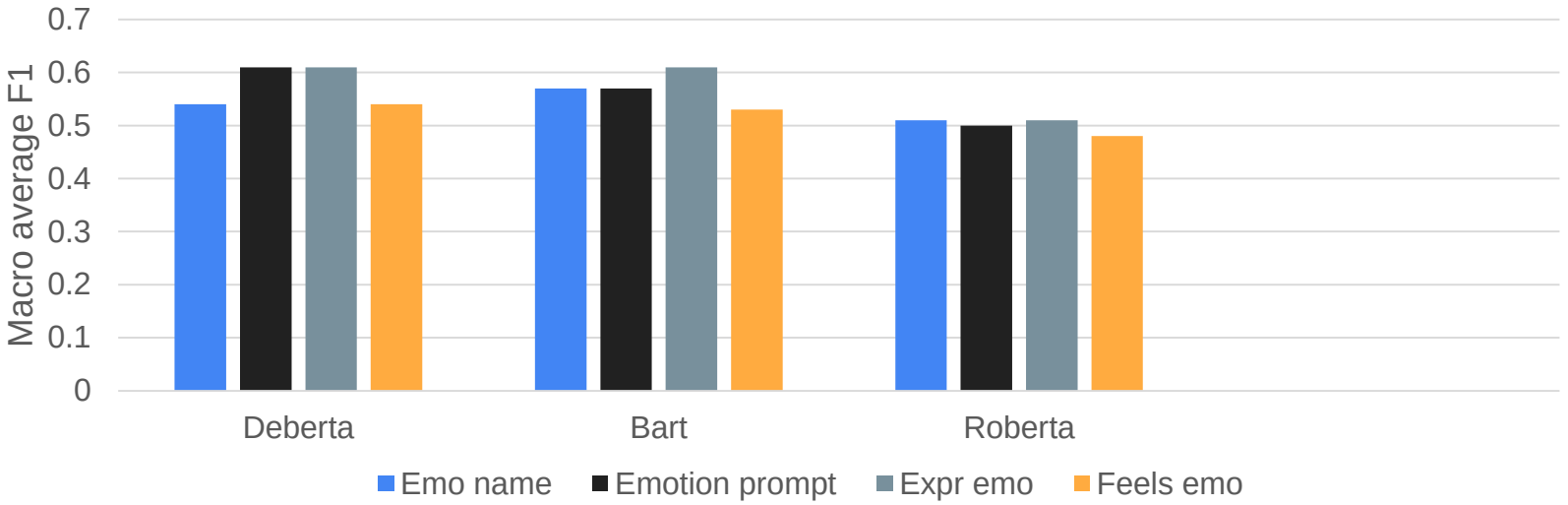


Figure 3: The results prompts of emotions (Isear dataset)

Results of prompts of emotions (Tec dataset)

When using Roberta model, the emotion prompt could lead to improvement

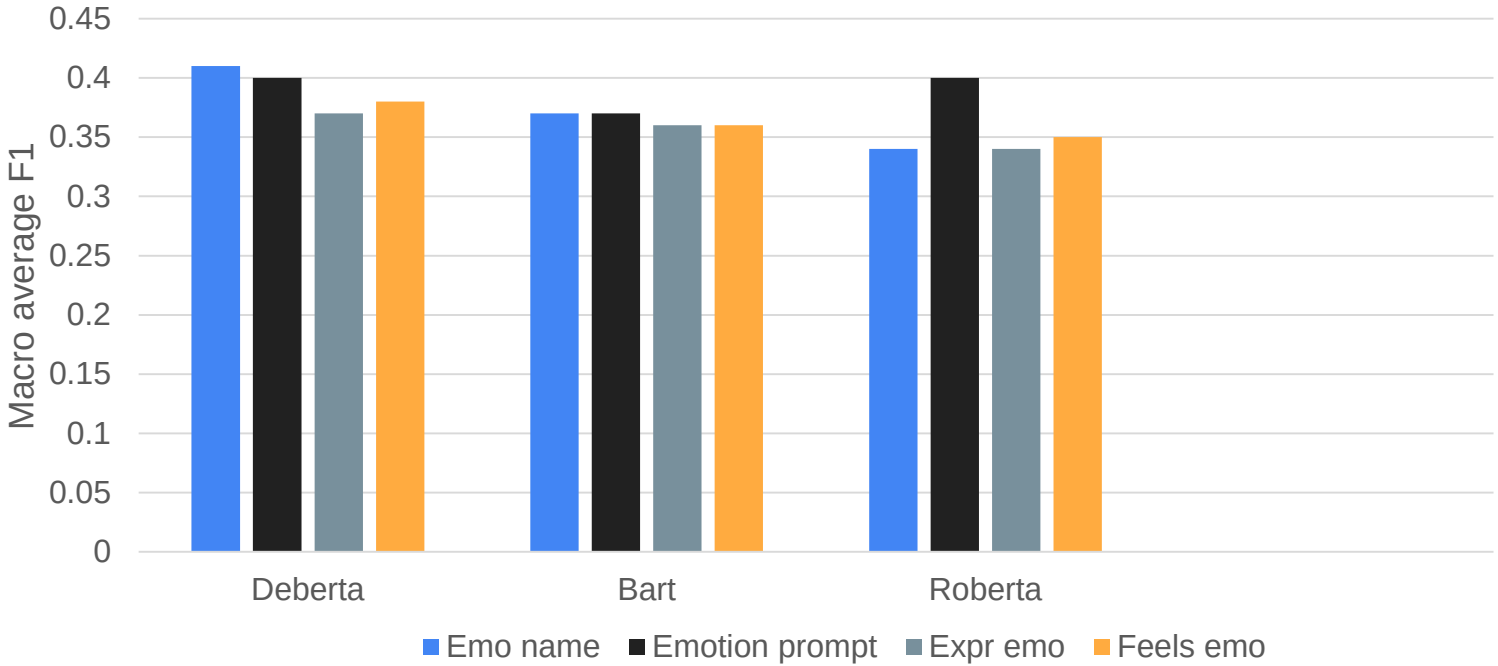


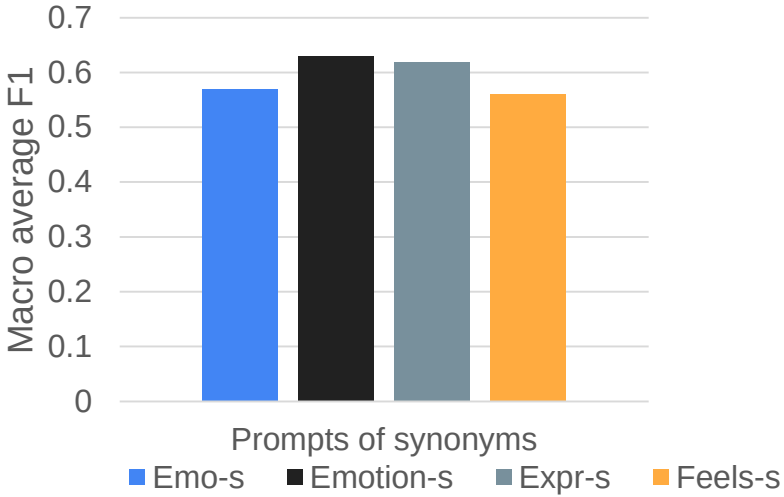
Figure 4: The results of prompts of emotions(Tec dataset)

.

The results of using synonyms of emotions (the average method, Isear dataset)

The experiments of using synonyms of emotions have been done using only Deberta model.

Using average of entailment probabilities of synonyms could lead to little improvement for all prompts for the Isear dataset



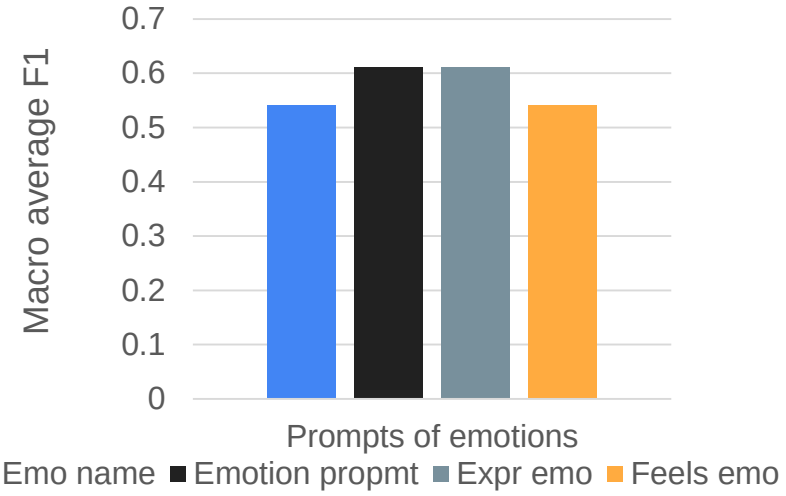
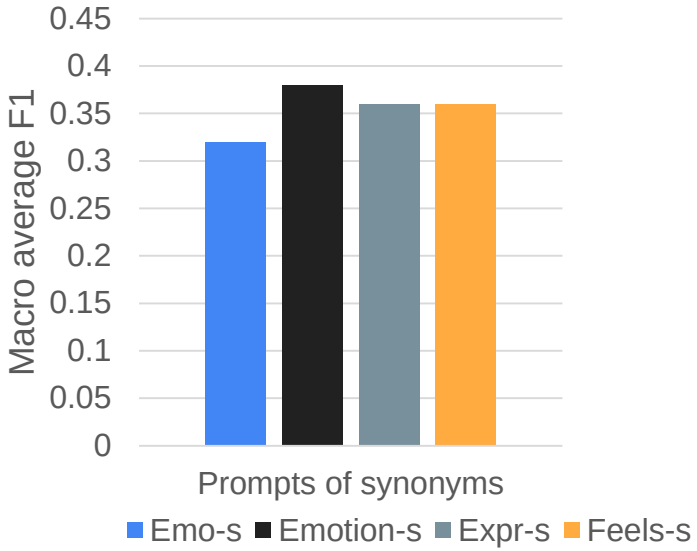


Figure 5: Comparison of the results of using synonyms and using prompts of emotions (The average method, Isear dataset)

The results of using synonyms of emotions (the average method, Tec dataset)

Using average of entailment probabilities of synonyms couldn't lead improvement for Tec dataset. (Using prompts of emotion names have higher performance.)



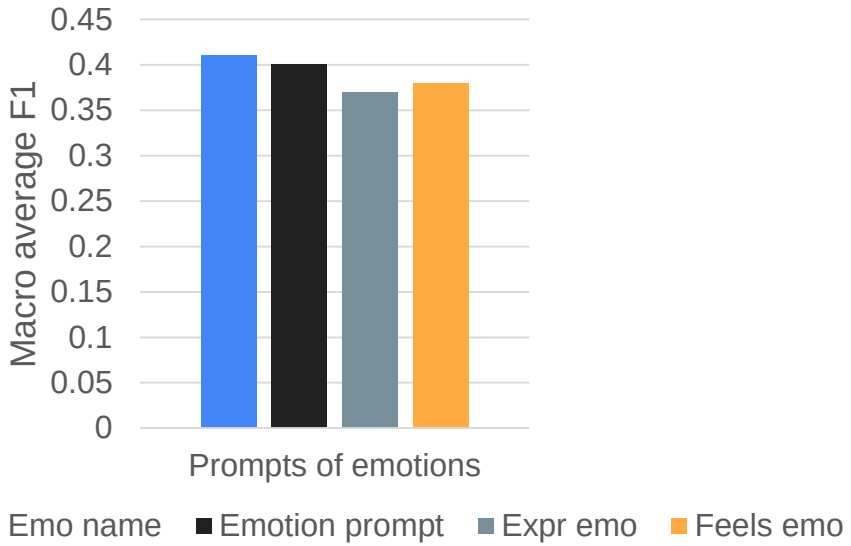


Figure 6: Comparison of the results of using synonyms and using prompts of emotions (The average method, Tec dataset)

The results of using synonyms of emotions(The maximum method, Isear dataset)

Using maximum of entailment probabilities of synonyms could lead to improvement just for feels_emo prompt compared to using prompt of only emotion name for the first dataset.

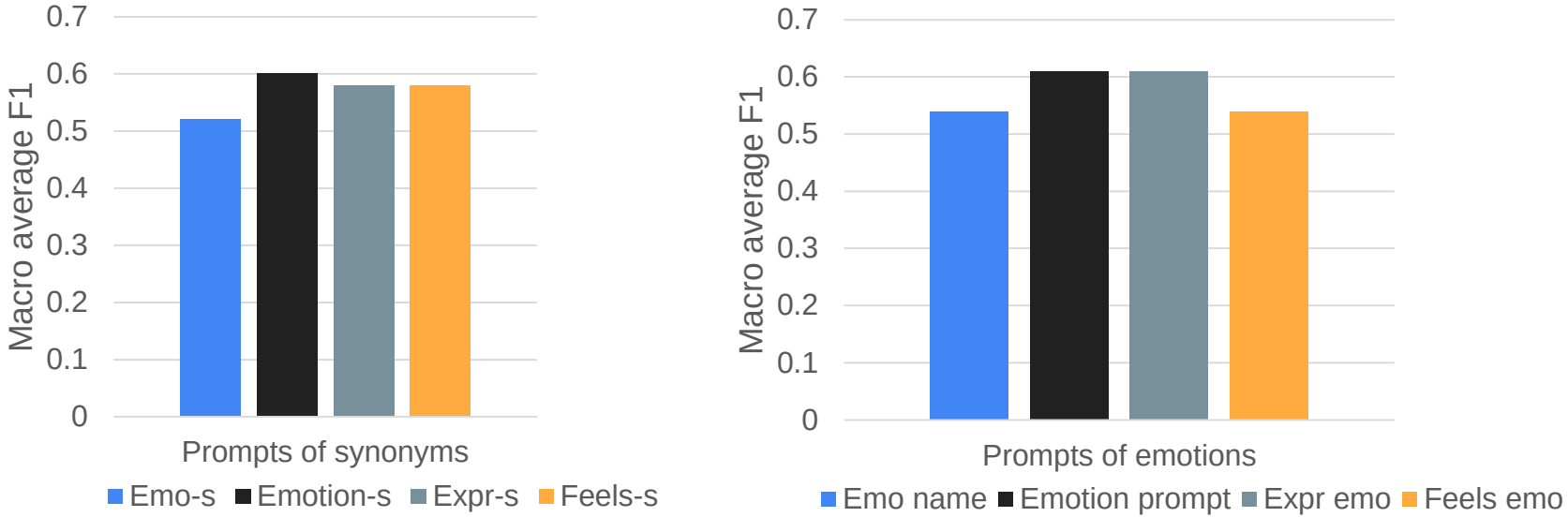
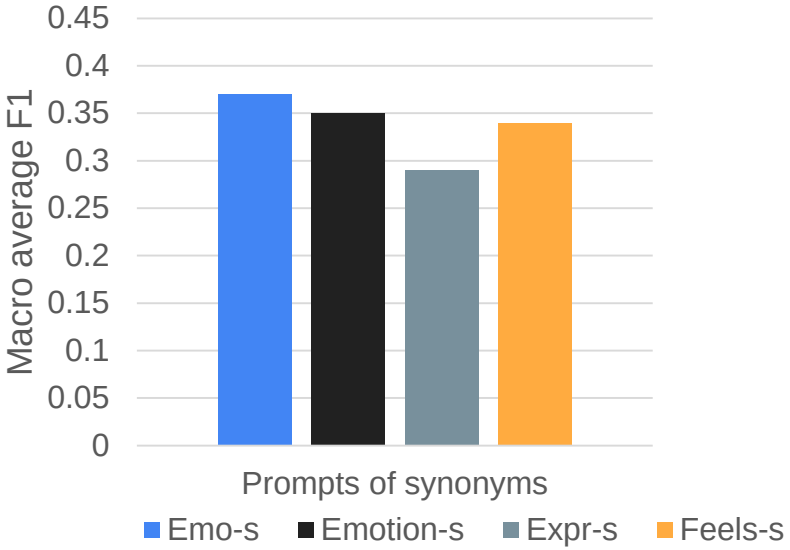


Figure 7: Comparison of the results of using synonyms and using prompts of emotions (The maximum method, Isear dataset)

The results of using synonyms of emotions (The maximum method, Tec dataset)

The prompts of emotions have better performance.



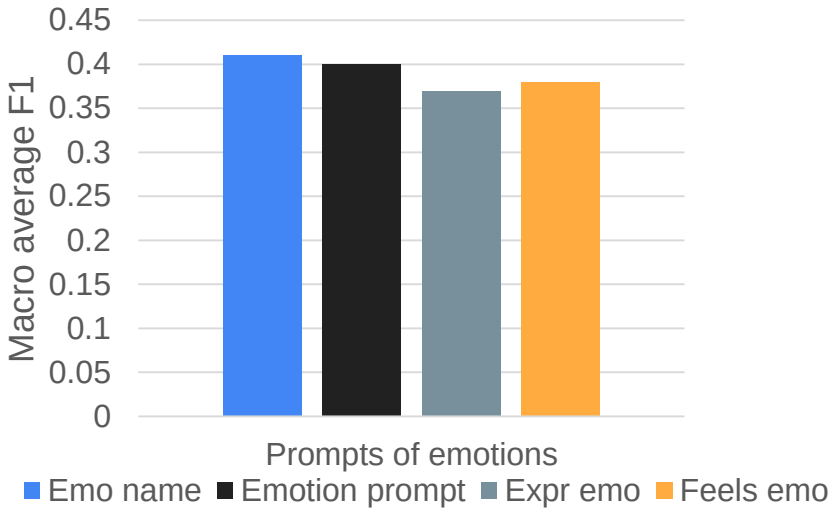


Figure 8: Comparison of the results of using synonyms and using emotion prompts (The maximum method, Tec dataset)

The result of combining two methods

Since emotion-s prompt had better performance compared to other prompts, the results of two previous methods for this prompt have been combined by getting average of entailment probabilities of each emotion.

- Isear dataset: Macro average F1: 0.64, accuracy:0.65
- Tec dataset: Macro average F1: 0.37, accuracy:0.41

it could lead to a little improvement for the first dataset

The effects of prompts on context of emotions

A sample sentences to compare the entailment probability of emotions with the text before and after using prompts:
Sentence: I heard part of a conversation in which one talked very low about women. (from Isear dataset)

When using prompt (adding “This text expresses” to the emotion) and Deberta model:

- Before using prompt: entailment probability of the true emotion disgust: 0.66

  entailment probability of the predicted emotion sadness: 0.76

- After using prompt: entailment probability of the true emotion disgust: 0.83

  The predicted emotion is true.

Comparing the prediction for some sample sentences

We can see that there are some specific tokens in the sentences that are related to the true labels. (like hurt related to anger, frustrating related to anger), but prediction by feels_emo prompt can also be reasonable

Table 4: Some sample sentences that using expr_emo could lead to improvement but feels_emo could not

True label	text	Emo name	Expr emo	Feels emo
anger	When friends try to put me down or hurt me. (Isear dataset)	fear	anger	fear
anger	It's really frustrating when your professors don't email back or show up for their office hours. (Tec dataset)	disgust	anger	fear

Comparing the prediction for some sample sentences

We can see that there are some tokens that are obviously related to the true emotions. (like death related to fear, cheer related to joy).

Table 5: Some sample sentences that both of expr_emo and feels_emo could lead to true prediction to them.

True label	Text	Expr emo	Feels emo
fear	Every time I imagine that someone I love or I could contact a serious illness, even death. (from Isear dataset)	fear	fear
joy	the moment when you get another follower and you cheer.(from Tec dataset)	joy	joy

Comparing the entailment probability for some sample sentences

The entailment probability for some sample sentences that their predictions are wrong.

When using average of entailment probabilities of synonyms

- When my partner was attacked and lost three teeth. (from Isear dataset, emotion-s prompt)

  True Label: anger, entailment probability: 0.84, predicted label: fear, entailment probability: 0.90

When using maximum of entailment probabilities of synonyms:

- In a cottage in a large forest, I was alone for a while in the dark. (from Isear dataset, emotion-s prompt)

  True label: fear, entailment probability: 0.92, predicted label: sadness, entailment probability: 0.98

These sentences have ambiguity and the entailment probabilities of true and predicted labels are close to each other.

The average of entailment probabilities for each emotion class (Isear dataset)

To compare the range of entailment probabilities for the cases that prediction is true and wrong, the average of entailment probabilities for each emotion class have been obtained

Table 6: The average of entailment probabilities for each emotion class (Isear dataset, when using average of entailment probabilities of synonyms, emotion-s prompt)

emotion	anger	disgust	fear	guilt	joy	sadness	shame
average (true prediction)	0.87	0.9	0.9	0.86	0.83	0.88	0.91
average (Wrong prediction)	0.71	0.62	0.68	0.75	0.35	0.6	0.45

The average of entailment probabilities for each emotion class (Tec dataset)

We can see that the range of entailment probabilities for true predictions are higher.

Table 7: The average of entailment probabilities for each emotion class (Tec dataset, when using average of entailment probabilities of synonyms, emotion-s prompt)

emotion	anger	Disgust	fear	Joy	sadness	surprise
average (true prediction)	0.84	0.86	0.61	0.75	0.79	0.77
average (wrong prediction)	0.47	0.52	0.31	0.23	0.39	0.57

Conclusion and future work

Conclusion

- Using prompts can affect the context of emotions and sometimes lead to improvement

- The synonyms of emotions have been used by different strategies, we saw that involving all of synonyms could lead to improvement

- The entailment probabilities of true emotion and predicted emotion are close to each other for the sentences that have ambiguity in their content

Limitation and future work

Limitations

- The prompts that have been used in this work were based on sample prompt types in recent studies
- The texts of datasets that have been used in these experiments had different complexities and some of them had ambiguity

Future work

- Designing a more guided prompt based on characteristics of datasets
- If the datasets with more consistent textual features were used, the synonym adaptive methods could lead to more improvement.
- Looking for a more guided synonym adaptive method

Thank you for listening