Emotion classification on code-mixed text messages via soft prompt tuning

Jinghui Zhang[1,*] and **Dongming Yang**[1,*,†] and **Siyu Bao**[1] and **Lina Cao**[1] and **Shunguo Fan**[1]
[1]China Telecom Cloud Technology Co., Ltd
{zhangjh33,yangdm1,baosy,caoln, fanshg}@chinatelecom.cn
† Corresponding author

Abstract

Emotion classification on code-mixed text messages is challenging due to the multilingual languages and non-literal cues (i.e., emoticons). To solve these problems, we propose an innovative soft prompt tuning method, which is lightweight and effective to release potential abilities of the pre-trained language models and improve the classification results. Firstly, we transform emoticons into textual information to utilize their rich emotional information. Then, variety of innovative templates and verbalizers are applied to promote emotion classification. Extensive experiments show that transforming emoticons and employing prompt tuning both benefit the performance. Finally, as a part of WASSA 2023, we obtain the accuracy of 0.972 in track MLEC and 0.892 in track MCEC, yielding the second place in both two tracks.

1 Introduction

Emotion plays an important role in social relationships, decision making, etc. Emotion analysis aims to enable machines to learn the emotions contained in textual information, such as conversations, e-commerce reviews and personal blogs (Balabantaray et al., 2012). With the surge of social media, emotion classification receives increasing attentions and brings huge commercial and social implications. [1] Text message is a common form of communication that exists in Twitter, YouTube, etc. Some text messages may contain typos, e.g., "habe" (have), and emotional words, e.g., "hahhhhh", while others may include visual cues such as icons, e.g., ":-)", and emoji. These features bring rich emotion information, but are typically not included in the vocabulary of pre-trained model like BERT (Bidirectional Encoder Representation from Transformers), T5 (Text-to-Text Transfer Transformer)(Raffel et al., 2020), etc. In addition, code-mixing is a critical challenge in emotion classification, which means that two or more languages (Yulianti et al., 2021), such as English and Roman Urdu, are contained in a single piece of text. Although code-mixed text is widely used, research on emotion classification for code-mixed corpus is still scarce (Ameer et al., 2022) since it is more difficult to recognizing emotions in code-mixing languages than in a monolingual language.

Fine-tuning is a widely-used method to improve performance of pre-trained language models (PLMs), however, it is memory-consuming and time-consuming. Compared to fine-tuning, prompt learning is more lightweight, thus has received increasing attentions (Liu et al., 2021a). T5 introduces a unified framework that converts any language problem into a text-to-text format. By introducing different prompts, the model can then adapt to different tasks, e.g., language translation and emotion classification. GPT-3 (General Pre-trained Transformer-3) (Brown et al., 2020) proposes context learning, where models can be applied directly in zero-shot or few-shot tasks without further fine-tuning. Utilizing language prompts helps to probe knowledge in PLMs and obtains better performance (Gu et al., 2021). Prompt learning has been widely applied in tasks of natural language understanding (Zhu et al., 2022; Yeh et al., 2022) and generation (Zheng and Huang, 2021), and even has been adopted in some vision tasks (Kirillov et al., 2023) and multi-modality tasks (Yao et al., 2021).

In this paper, we present our effort in emotion classification on code-mixed text messages. The contributions are summarized as follows:

- A concise emoticon pre-processing method is proposed to transform emoticons into textual information.
- Variety of innovative templates are designed to improve emotion classification on code-mixed text messages.
- Specific verbalizers are applied, which

[1]* These authors contributed equally to this work.

Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis, pages 596–600
July 14, 2023

achieves efficient emotion classification with low resource consumption.

2 Related Work

2.1 Emotion Classification

Because of the vagueness of definition and the similarity of some emotions, emotion classification is often seen as a subjective and challenging task.

Gaind et al. (2019) combines two different approaches to extract emotions. The first approach employs several textual features like emoticons, degree words, negations and other grammatical analysis. The second approach adopts algorithms based on machine learning techniques.

Polignano et al. (2019) proposes a classification method based on deep neural networks, which is effective on different emotion classification datasets. They also compare three pre-trained word embeddings for words encoding.

Hasan et al. (2019) uses a dimensional model to define emotion categories. Besides, they propose a soft classification method to measure the probability of assigning messages to each emotion category. In addition, a framework called EmotexStream is developed for real-time emotion tracking.

Ameer et al. (2023) proposes multiple attention mechanisms that reveal the contribution of each word to each emotion. They also investigate the usage of LSTM, the fine-tuning of transformer networks for multi-label emotion classification. Experimental results show that these new transfer learning models are able to outperform current state-of-the-arts on the SemEval-2018 Task-1C dataset.

Ashraf et al. (2022) creates the first multi-label emotion dataset, consisting of six basic emotions from the Urdu Nastalíq script. In addition, they build a set of baseline classifiers and provide insight into these classifiers. The baseline classifiers include machine learning algorithms (i.e., Random Forest, Decision Tree, Sequential Minimal Optimisation, AdaBoostM1 and Bagging), deep learning algorithms and Baseline Based Transformer (i.e., BERT). They use a combination of text representations, which are stylometric-based features, pre-trained word embedding, word-based n-grams, and character-based n-grams.

2.2 Prompt Learning

Prompt learning typically includes the design of prompt templates, the optimization of verbalizers and the selection of the PLM. Prompt templates and verbalizers are usually produced by expert knowledge or generated automatically using search or optimization methods (Hu et al., 2021; Shin et al., 2020; Gao et al., 2021; Liu et al., 2021a). The selection of PLMs needs to take the model structure(such as auto-regressive model), pre-training dataset into consideration.

Liu et al. (2021a) finds that prompt tuning can be effective among different models and natural language tasks. They propose a universal and simple P-tuning v2 method, which proves that prompt tuning can be comparable to fine-tuning, while only 0.1%-3% of the parameters are fine-tuned.

Hambardzumyan et al. (2021) proposes soft verbalizers for prompt learning. They use a continuous vector for each class and adopt dot multiplication between the output of masked language model and the class vector to generate the probabilities.

Lang et al. (2022) finds that co-training can improve the prompt-based learning using unlabelled data. Specifically, co-training can benefit the original prompt model while learning smaller downstream task-specific models.

Han et al. (2022) proposes a prompt tuning method with rules "PTR", which encodes the prior knowledge of a classification task into a rule. They then design sub-prompts based on the rule to adapt the task. Results show that PTR achieves a good trade-off between effectiveness and efficiency in prompt construction compared to the state-of-arts.

In this paper, we propose innovative prompt tuning techniques and bring valuable attempts for emotion classification on code-mixed text messages.

3 Methods

3.1 Emoticon Pre-processing

Text messages may include icons and Unicode emoji, which provide rich emotional information and are overlooked by many researchers. We propose a concise method to convert these features into textual form, which consists of icon-emotion mapping and Unicode-short name mapping.

We first statistically gather the icons in the training set and collect more icons from the Internet as a supplement. An icon-emotion mapping is then constructed based on the visual features of icons. For example, ":-)" is replaced by [joy]. It is worth mentioning that some icons can represent different emotions in different contexts. For example, ";-/" can indicate emotions such as disgust, anger and sadness. In this case, ";-/" is replaced by [disgust,

ID	Template
Manual 1	A text message of [mask]: [x]
Manual 1F	A text message of joy: Good job! A text message of [mask]: [x]
Manual 2	[x] This talk is about [mask].
Manual 2F	Good job! This talk is about joy. [x] This talk is about [mask].

Table 1: Manual templates based on expert knowledge. F: few-shot prompt, [x]: input text message.

Method	MLEC	MCEC
MT+MV	0.024	0.842
MT+MV+EP	0.082	0.856
ST+SV	0.879	0.966
ST+SV+EP	**0.892**	**0.972**

Table 2: Accuracy on test sets for both tracks. M: manual, S: soft, T: template, V: verbalizer, EP: emoticon pre-processing.

anger, sadness], informing the PLM that this text may contain some of these three feelings. Since only text replacements are performed, the computation cost of this operation is trivial.

For Unicode-short name mapping, we convert the Unicode encoding to a CLDR short name, e.g. replacing U+1F600 with [grinning face].

3.2 Prompt Tuning

According to the experience (Liu et al., 2021b), a single word change in prompts could make a drastic difference. Therefore, we tried a variety of prompts.

Firstly, we design textual prompts following the zero-shot and few-shot approach of GPT-3, as shown in Table 1. Considering the short length of the text messages in CM-MEC-21 corpus, we limit the number of manual prompt samples for the few-shot approach.

Then we design several soft prompts following Zhu et al. (2022). We investigate the effect of soft template with different numbers and distribution of soft tokens. The structure of soft template is shown in Fig 1. Finally, we compare the results of the manual verbalizer and the soft verbalizer (Hambardzumyan et al., 2021).

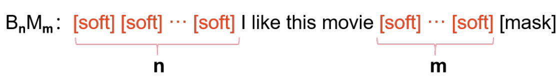

Figure 1: Soft template with different numbers and distribution of soft tokens. B: beginning, M: middle. The *n* and *m* denote the numbers of tokens.

For the PLMs, we adopt XLM-RoBERTa with different model scales and pre-training sets. In the process of optimizing the templates and verbalizers, we just freeze the parameters of the pre-trained model, as shown in Fig 2.

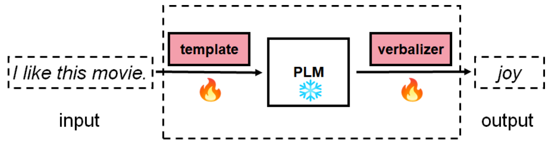

Figure 2: Configuration of prompted model.

Ultimately, the equations for result prediction and objective optimization are noted as

$$\hat{p}_{\nu,\tau}\left(\mathrm{x}\right) = \mathrm{softmax}\, V_{\nu}\left(F\left(T_{\tau}\left(x\right)\right)\right),$$

$$\{v^{*}, \tau^{*}\} = \arg\max_{\nu,\tau} \sum p_{\nu,\tau}\left(x\right) \log \hat{p}_{\nu,\tau}\left(x\right),$$

where τ, ν denote the parameters of the template and verbalizer respectively. T, F, V represent the template, PLM and verbalizer respectively. The x refers to the input text message, and $\hat{p}$ represents the estimated emotion probability distribution corresponding to the text message while p represents real distribution.

4 Results and Discussion

We adopt CM-MEC-21 corpus (Ameer et al., 2022) in this work, which includes 11650 MCEC messages and 11603 MLEC messages. We combine MCEC and MLEC text messages for both training and evaluating. The accuracy based on the merged dataset for both tracks are summurized in Table 2.

4.1 Emoticon Pre-processing

Firstly, we analyze the effects of emoticon pre-processing. We take XLM-RoBERTa fine-tuned on Roman Urdu corpus[2] as the baseline, where soft template B_5M_6 and manual verbalizer are used but emoticon pre-processing are not employed. As shown in Fig 3, both parts of emoticon pre-processing (i.e., icon-emotion mapping and Unicode-short name mapping) are effective in helping the machine to recognize the emotions.

[2] https://huggingface.co/Aimlab/xlm-roberta-roman-urdu-finetuned

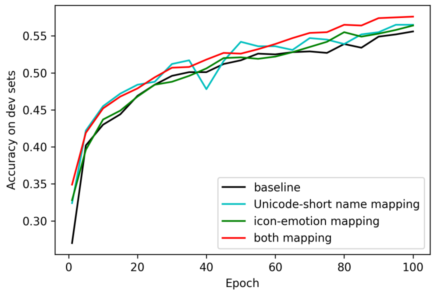

Figure 3: Comparison of different data pre-processing.

4.2 PLMs Comparison

We compare the performance of different PLMs, as shown in Fig 4. We use different XLM-RoBERTa models, where soft template B_5M_6, soft verbalizer and emoticon pre-processing are adopted. We can see that larger models bring better performance. Meanwhile, the language used during pre-training has a significant impact on the final results.

After about 25 epochs of training, XLM-RoBERTa-fine-tuned achieves an accuracy of 0.9, indicating that fine-tuned model can better understand the code-mixed messages. We can draw a conclusion from the experiments that, with the fine-tuned model, just a small amount of prompt tuning can achieve satisfactory results. By contrast, more prompt tuning is required while employing general pre-trained models.

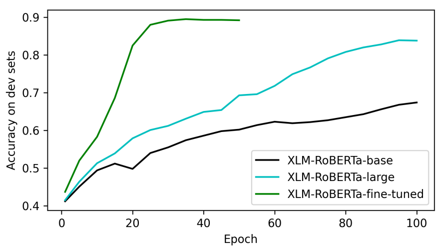

Figure 4: Comparison of different PLMs.

4.3 Prompt Tuning

The combination of template and verbalizer significantly affects classification performance, as shown in Table 3. Although manual prompts are relatively easier to design, their performances are far inferior to that of soft prompts.

We can see from Table 3 that a well-designed prompt template is a prerequisite for efficient classification. Otherwise, the effect of the prompt will be limited, such as Manual 1 and Manual 2F. Also, increasing soft tokens in the template can improve the result, but only to a limited extent.

Template	ID	MV	SV
manual	Manual 1	0.043	0.353
	Manual 1F	0.353	0.353
	Manual 2	0.107	0.887
	Manual 2F	0.353	0.559
soft	B_0M_1	0.434	0.892
	B_0M_3	0.49	0.894
	B_0M_5	0.503	0.896
	B_0M_6	0.551	0.897
	B_1M_6	0.556	0.895
	B_3M_6	0.556	0.892
	$\mathbf{B_5M_6}$	0.576	**0.902**

Table 3: Comparison of different prompt tuning. MV: manual verbalizer, SV: soft verbalizer.

Soft verbalizer can universally benefit the classification performance. For example, while employing template of Manual 2, the soft verbalizer brings an accuracy gain of 0.780 compared with the manual verbalizer. Besides, soft verbalizer enables low resource consumption while maintaining satisfactory performance, e.g., just one soft token in the template can achieve an accuracy of 0.9.

4.4 Limitations

While dealing with the MLEC task, we find that most outputs from our method have only one category. Therefore, our method still has room for improvement in handling MLEC problem.

5 Conclusion

In this paper, we propose the innovative prompt tuning for emotion classification on code-mixed text messages. We first transform emoticons into textual information to utilize their rich emotional information. Then, variety of innovative templates and verbalizers are applied to promote emotion classification. The results show that transforming emoticons benefits the final results. More importantly, even without model fine-tuning, optimizing the prompts yields efficient classification. Finally, we obtain the accuracy of 0.972 in track MLEC and 0.892 in track MCEC, yielding the second place in both two tracks. In future work, we will pay more attention to combining model fine-tuning with prompt learning, and further explore the capacity of PLMs on emotion classification.

References

Iqra Ameer, Necva Bölücü, Muhammad Hammad Fahim Siddiqui, Burcu Can, Grigori Sidorov, and Alexander Gelbukh. 2023. Multi-label emotion classification in texts using transfer learning. *Expert Systems with Applications*, 213:118534.

Iqra Ameer, Grigori Sidorov, Helena Gomez-Adorno, and Rao Muhammad Adeel Nawab. 2022. Multi-label emotion classification on code-mixed text: Data and methods. *IEEE Access*, 10:8779–8789.

Noman Ashraf, Lal Khan, Sabur Butt, Hsien-Tsung Chang, Grigori Sidorov, and Alexander Gelbukh. 2022. Multi-label emotion classification of urdu tweets. *PeerJ Computer Science*, 8:e896.

Rakesh C Balabantaray, Mudasir Mohammad, and Nibha Sharma. 2012. Multi-class twitter emotion classification: A new approach. *International Journal of Applied Information Systems*, 4(1):48–53.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

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

Tianyu Gao, Adam Fisch, and Danqi Chen. 2021. Making pre-trained language models better few-shot learners. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3816–3830, Online. Association for Computational Linguistics.

Yuxian Gu, Xu Han, Zhiyuan Liu, and Minlie Huang. 2021. Ppt: Pre-trained prompt tuning for few-shot learning. *arXiv preprint arXiv:2109.04332*.

Karen Hambardzumyan, Hrant Khachatrian, and Jonathan May. 2021. Warp: Word-level adversarial reprogramming. *arXiv preprint arXiv:2101.00121*.

Xu Han, Weilin Zhao, Ning Ding, Zhiyuan Liu, and Maosong Sun. 2022. Ptr: Prompt tuning with rules for text classification. *AI Open*, 3:182–192.

Maryam Hasan, Elke Rundensteiner, and Emmanuel Agu. 2019. Automatic emotion detection in text streams by analyzing twitter data. *International Journal of Data Science and Analytics*, 7:35–51.

Shengding Hu, Ning Ding, Huadong Wang, Zhiyuan Liu, Jingang Wang, Juanzi Li, Wei Wu, and Maosong Sun. 2021. Knowledgeable prompt-tuning: Incorporating knowledge into prompt verbalizer for text classification. *arXiv preprint arXiv:2108.02035*.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. 2023. Segment anything. *arXiv preprint arXiv:2304.02643*.

Hunter Lang, Monica N Agrawal, Yoon Kim, and David Sontag. 2022. Co-training improves prompt-based learning for large language models. In *International Conference on Machine Learning*, pages 11985–12003. PMLR.

Xiao Liu, Kaixuan Ji, Yicheng Fu, Weng Lam Tam, Zhengxiao Du, Zhilin Yang, and Jie Tang. 2021a. P-tuning v2: Prompt tuning can be comparable to fine-tuning universally across scales and tasks. *arXiv preprint arXiv:2110.07602*.

Xiao Liu, Yanan Zheng, Zhengxiao Du, Ming Ding, Yujie Qian, Zhilin Yang, and Jie Tang. 2021b. Gpt understands, too. *arXiv preprint arXiv:2103.10385*.

Marco Polignano, Pierpaolo Basile, Marco de Gemmis, and Giovanni Semeraro. 2019. A comparison of word-embeddings in emotion detection from text using bilstm, cnn and self-attention. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*, pages 63–68.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1):5485–5551.

Taylor Shin, Yasaman Razeghi, Robert L Logan IV, Eric Wallace, and Sameer Singh. 2020. Autoprompt: Eliciting knowledge from language models with automatically generated prompts. *arXiv preprint arXiv:2010.15980*.

Yuan Yao, Ao Zhang, Zhengyan Zhang, Zhiyuan Liu, Tat-Seng Chua, and Maosong Sun. 2021. Cpt: Colorful prompt tuning for pre-trained vision-language models. *arXiv preprint arXiv:2109.11797*.

Hui-Syuan Yeh, Thomas Lavergne, and Pierre Zweigenbaum. 2022. Decorate the examples: A simple method of prompt design for biomedical relation extraction. *arXiv preprint arXiv:2204.10360*.

Evi Yulianti, Ajmal Kurnia, Mirna Adriani, and Yoppy Setyo Duto. 2021. Normalisation of indonesian-english code-mixed text and its effect on emotion classification. *International Journal of Advanced Computer Science and Applications*, 12(11).

Chujie Zheng and Minlie Huang. 2021. Exploring prompt-based few-shot learning for grounded dialog generation. *arXiv preprint arXiv:2109.06513*.

Yi Zhu, Xinke Zhou, Jipeng Qiang, Yun Li, Yunhao Yuan, and Xindong Wu. 2022. Prompt-learning for short text classification. *arXiv preprint arXiv:2202.11345*.