The 6th International Conference on Interdisciplinary research on Computer science, Psychology, and Education (ICICPE' 2022)
December 26-28, 2022. Pattaya, Thailand.

Prompt-based Learning for English-German Critical Translation Error Detection

[1] Sugyeong Eo, [2] Chanjun Park, [1] Heuiseok Lim*

[1] Korea University [2] Upstage
{djtnrud, limhseok}@korea.ac.kr chanjun.park@upstage.ai

Abstract. Critical error detection refers to detecting cases in which machine translation errors cause deviations in the meaning of law, society, safety, numbers, etc. It is possible to prevent social repercussions and ethical problems, and studies are being conducted. This study exploits prompt-based learning to maximize the language understanding capability of the language model. As a result of experiments on the WMT (Conference on Machine Translation) 21 English-German critical error detection dataset, our methodology outperforms the WMT official baseline performance.

Keywords: prompt, neural machine translation, quality estimation, critical error detection

1 Introduction

Machine translation (MT) systems translate source sentences fluently, but there are cases where critical errors are included. A critical error refers to a case in which the meaning of the MT result is deviate with a statement that may cause economic, legal, safety and ethical, and social problems. In this case, translation errors cause serious social impact or loss. Therefore, a task called critical error detection (CED) was first introduced in WMT (Conference on Machine Translation) 21.

The CED is a binary classification task to classify only critical errors from correct and simple translation errors. We use the WMT21 official English-German critical error dataset and train the model by adopting prompt-based learning[1,2]. We use only the XLM-RoBERTa [3] model without using additional parallel data. Our approach outperforms the baseline performance, and we find that our methodology was effective.

* Corresponding Author

2 Prompting

We design intuitive hard prompts manually. Depending on the input sequence configuration, the performance of the model can vary greatly, and the effective prompt for the CED task is unknown. Therefore, we designed various prompt candidates. First, in prompt engineering that composes the template, we select description along with the source sentence and the translation sentence. In the case of a verbalizer that maps the actual label and the word candidate for the masked token, we engineer words with clear contrasts.

3 Experiments

We use the English-German critical error detection dataset provided by WMT21. As a metric for evaluation, we use Matthew's correlation coefficient (MCC). We used XLM-RoBERTa large for task learning, and leverage the model and tokenizer distributed by Huggingface. When performing prompt-based fine-tuning, we experiment with the LM-BFF framework.

Table1: Results for the WMT21 English-German critical error detection dev and test set.

Template	Dev MCC	Test Mcc
Input + [great/terrible] translation	0.46	0.47
Input + [good/bad] translation	0.49	0.49
Input + [!/?]	0.45	0.45
Baseline	-	0.40

Table 1 indicates the results of critical error detection for English-German. Our approach overall outperforms the baseline performance. There is a big difference in performance for each prompt and verbalizer, and the experiment using the description of "Input + [good/bad] translation" shows the highest performance.

4 Conclusion

We performed a critical error detection task to find critical errors that cause semantic deviations in the translation process. We train the XLM-RoBERTa model by adopting prompt-based learning. Our method is simple yet powerful, and we have proven its effectiveness in experiments.

5 Acknowledgement

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2022-2018-0-01405) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation)

References

1. Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2021a. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. arXiv preprint arXiv:2107.13586
2. Tianyu Gao, Adam Fisch, and Danqi Chen. 2020. Making pre-trained language models better few-shot learners. arXiv preprint arXiv:2012.15723
3. Conneau et al. (2019). Unsupervised cross-lingual representation learning at scale. arXiv preprint arXiv:1911.02116. DOI : 10.18653/v1/P19-4007