The 7th International Conference on Interdisciplinary Research
on Computer Science, Psychology, and Education (2023)

Enhancing Translation Performance with SuperICL4Gen

SeongTae Hong∘†, Seungjun Lee†, Sungmin Ahn§ HeuiSeok Lim*†‡
Department of Computer Science and Engineering, Korea University†, O2O Inc.§ Human-Inspired AI Research‡
{ghdchlwls123,dzzy6505,limhseok}@korea.ac.kr
smahn@o2o.kr

Abstract

At present, Large language models like GPT-4[1] are demonstrating exceptional performance across a range of tasks, including machine translation, summarization, and conversational applications. However, these Large language models come with their own set of challenges, such as the requirement for significant computational resources and the complexities of fine-tuning for specific domains during both training and deployment. While In-Context Learning has provided some efficiency by effectively operating with context information extracted from datasets and partially addressing these issues, it still faces challenges. To tackle these challenges, the concept of SuperICL has been introduced. SuperICL[2] involves restructuring context using the output information from the applied fine-tuned PLM model, enhancing the ability of the Large Language Model (LLM) to classify more effectively. SuperICL4Gen is a methodology specifically devised for generative tasks, and this research verify its effectiveness through its application in translation tasks.

Keywords: SuperICL4Gen, In-Context Learning, Translation

1. Introduction

The advent large-scale language models has heralded a new era of advancements in numerous natural language processing domains, including machine translation, summarization, and conversational tasks. OpenAI's GPT-4[1] shows human-level competence in tasks related to both understanding and generating natural language. Fine-tuning large language models to various domains can be challenging, especially when it requires specialized knowledge or when dealing with new tasks. In-Context Learning has emerged as a methodology for effectively optimizing Large Language Models (LLMs) using contextual information from available data to address existing tasks. However, it is sensitive to both the quantity and sequence of context shots provided. In response to this limitation, SuperICL has emerged as a methodology to provide additional information to LLM, improving its robustness and performance. This research presents a methodology employing SuperICL[2] for translation tasks. Through this approach, exceptional performance is observed in translation tasks. Using evaluation metrics such as BLEU[3] and ROUGE[4], we compare the results of translation with Super4Gen to those of existing models. In these evaluation metrics, translation results using Super4Gen significantly outperform the existing models.

2. Methods

2.1 Fine-Tuning of Plugin Models

In this research, We combine a fine-tuned generative model, trained using supervised learning datasets, with a massive language model. The fine-tuned model provides the specific knowledge and capabilities needed for translation tasks in the target language and collaborates to improve overall performance.

2.2 Input Modification for LLM

Existing In-Context Learning primarily uses context related to specific tasks as input, typically composed of examples sampled from datasets. These examples include input text along with their corresponding answers, forming the context used for each task. In SuperICL, the generated results from the fine-tuned model are reintroduced as additional information into the input of the Large Language Model (LLM). Unlike SuperICL, where the probability values of classification labels are included as context informa-

*Corresponding author

tion, SuperICL4Gen utilizes a fine-tuned model to generate translated predicted sentences for the input. These predicted translation sentences are used to calculate the average values of token generation probabilities, which are then employed as the final probability values. This process involves utilizing shots randomly sampled from the training dataset, encompassing various contexts and examples, enabling the model to operate and adapt to diverse situations. Each shot includes the following information:

- Input Sentence: The original input sentence for translation tasks.
- Generated Sentence: The sentence translated by the fine-tuned model.
- Generation Probability: The probability of the fine-tuned model's prediction outcomes.
- Answer: The correct translation of the input sentence.

Similarly, for test , fine-tuned model is used to reconstruct the input given the output and probabilities for the input sentence. Through this approach, In SuperICL4Gen, the Large Language Model generates the final predictions by considering the translation results and probabilities from the fine-tuned model.

3. Experiments

In this experiment, the KE-T5 Base[5] model serves as the fine-tuned model. Data is sourced from the AIHub's daily life and spoken language Korean-English parallel corpus[1], with a training dataset consisting of 121,696 examples and a validation dataset of 34,770 examples. To assess the models, a test dataset of 100 samples is employed, and for each model, three random shots are provided as context. Each model receives the same set of provided shots.

table 1 presents the translation performance results between English and Korean, as well as Korean and English. The model utilizing SuperICL4Gen achieves the highest BLEU score (38.20) and ROUGE-L score (0.7357), demonstrating superior performance in English to Korean translation. Additionally, in the translation from English to Korean, it achieves a BLEU score (37.30) and ROUGE-L score (0.7260), indicating an overall improvement in performance.

[1] https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=71265

Table 1. Quantitative evaluation results and performance comparison

model	BLEU	ROUGE-1	ROUGE-2	ROUGE-L
English-Korean				
KE-T5-base	34.12	0.7434	0.6149	0.7023
GPT-3.5 ICL	21.79.	0.7634	0.6378	0.6972
SuperICL4Gen	**37.30**	**0.7766**	**0.6536**	**0.7260**
Korean-English				
KE-T5-base	28.40	0.7605	0.6544	0.6604
GPT-3.5 ICL	35.07	**0.8634**	0.7511	0.7225
SuperICL4Gen	**38.20**	0.8601	**0.7530**	**0.7357**

4. Conclusion

SuperICL4Gen demonstrate outstanding performance in translation tasks. It significantly outperform the existing baseline and the model adopting In-Context Learning methodology in the quantitative evaluation metrics identified in the experimental results. In conclusion, this research opens up new possibilities for translation tasks and validates the utility of integrating existing PLMs into LLM, thereby providing a methodology for performing supervised learning with conventional PLM models.

Acknowledgements

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). This work was supported by Institute of Information communications Technology Planning Evaluation(IITP) grant funded by the Korea government(MSIT) (No. 2020-0-00368, A Neural-Symbolic Model for Knowledge Acquisition and Inference Techniques). Following are results of a study on the "Leaders in INdustry-university Cooperation 3.0" Project, supported by the Ministry of Education and National Research Foundation of Korea.

Reference

[1] OpenAI, "Gpt-4 technical report," 2023.

[2] C. Xu, Y. Xu, S. Wang, Y. Liu, C. Zhu, and J. McAuley, "Small models are valuable plug-ins for large language models," *arXiv preprint arXiv:2305.08848*, 2023.

[3] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation,"

Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pp. 311–318, Jul. 2002. [Online]. Available: https://aclanthology.org/P02-1040

[4] C.-Y. Lin, "ROUGE: A package for automatic evaluation of summaries," *Text Summarization Branches Out*, pp. 74–81, Jul. 2004. [Online]. Available: https://aclanthology.org/W04-1013

[5] K. AIRC, "Ke-t5: Korean english t5," Mar. 2021. [Online]. Available: https://github.com/AIRC-KETI/ke-t5