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

A Comparative Study on Cross-Lingual Post-Training (XPT) with Korean

Suhyune Son[1], and Aram So[2*]

[1] Department of Computer Science, Korea University
[2] Human-inspired AI Research
{ssh5131, aram}@korea.ac.kr

Abstract. There is a limitation to constructing a large corpus for training a pre-trained language model in low resource languages. In this paper, we apply the Cross-lingual Post-Training (XPT) method, which overcomes these limitations, and analyze the effectiveness of the method in Korean, which has few resources. We conduct overall performance comparison and analysis studies with Korean pre-trained model and mBERT using only a small amount of Korean corpus, 4M.

Keywords: Transfer Learning, Korean Language Model, Cross-lingual Transfer Learning

1 Introduction

Recently, studies on various natural language processing tasks have been actively conducted with large amounts of data. Despite of this necessity and importance, the language model studies are based on English. Since the amount of training data is a factor directly related to the performance of a language model, an imbalance of language resources may become an obstacle to performance improvement. To alleviate this limitation, we construct an efficient Korean language model by applying Cross-lingual Post-Training (XPT) [1] to the English language model. In this paper, we prove the effectiveness of the XPT method by comparison with various Korean language models based on KLUE [2], a Korean benchmark dataset.

2 Experiments

XPT performs transfer learning between two languages by using a small amount of Korean corpus. In Phase 1, after fixing all parameters of the encoder layer of the source language model, only the Implicit Translation Layer (ITL) that is additionally added to the input and output of the embedding layer and the encoding layer is trained. In Phase 2, we train all embeddings, encodings, outputs, and ITLs.

*Corresponding Author

Table 1. The comparison results of the XPT model with the Korean/multi-lingual language model.

	YNAT	STS	NLI	NER	RE	DP	MRC
	F1	F1	ACC	F1	F1	UAS	ROUGE
KoELECTRA	85.28	84.77	82.37	66.35	60.62	93.85	61.58
mBERT	82.14	78.76	71.45	76.12	57.25	90.45	55.56
XPT	86.25	89.36	80.47	71.02	61.78	91.45	30.2

As in Table 1, XPT-4M, trained with a Korean corpus of 4M, showed the best or second-best performance in all tasks except KLUE-NLI and KLUE-MRC compared to KoELECTRA† and mBERT [3]. Considering that training was performed using a significantly smaller amount of corpus compared to the previous study, it can be said that it is meaningful just to show similar performance.

3 Conclusion

In this paper, we use the Cross-lingual Post-Training (XPT) method that can improve the performance of the model by utilizing the source language pre-trained model in low resource language. As a result, XPT showed better performance than the existing language model trained with a large corpus in most tasks. This shows the necessity and effectiveness of the XPT methodology in low language.

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 the Technology Innovation Program (or Industrial Strategic Technology Development Program) (20014847, Development of consumer-customized live commerce and untouch order technology for senior) funded By the Ministry of Trade, Industry & Energy(MOTIE, Korea)

References

1. Lee, C., Yang, K., Whang, T., Park, C., Matteson, A., & Lim, H. (2021). Exploring the data efficiency of cross-lingual post-training in pretrained language models. Applied Sciences, 11(5), 1974.
2. Park, S., Moon, J., Kim, S., Cho, W. I., Han, J., Park, J., ... & Cho, K. (2021). Klue: Korean language understanding evaluation. arXiv preprint arXiv:2105.09680.
3. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

† https://github.com/monologg/KoELECTRA