Examining the Transformer Models with the K-CoLA Dataset

Yong-hun Lee
(Chungnam National University)

Since the introduction of the Transformer model (Vaswani et al., 2017), there have been many trials to apply deep learning technology to study natural languages, especially after the advent of the Bidirectional Encoder Representations from Transformers (BERT; Devlin et al., 2019). Warstadt et al. (2019) also provided the Corpus of Linguistic Acceptability (CoLA) dataset, by which it was possible to explore some syntactic phenomena in English.

However, not all the syntactic phenomena can effectively and correctly be analyzed with the Transformer models. Warstadt et al. (2019), for example, classified their dataset into 8 groups (*simple* labels sentences with no marked syntactic structures; *adjunct* labels sentences that contain adjuncts of nouns and verb phrases; *comp* clause labels sentences with embedded or complement clauses*; to-VP* labels sentences with non-finite embedded verb phrase; *arg altern* labels sentences with non-canonical argument structures such as passives; binding labels sentences with pronouns and binding phenomena; *question* labels sentences with interrogative clauses and relative clauses; and *violations* labels sentences with morphological or semantic violations, or an extra/missing word) and examined their performances. They found that each phenomenon demonstrated different performances.

This study applied expanded the approach to the Korean data and tried to examine the performance of three different BERT models. They were all based on the BERT (Devlin et al., 2019). For the comparisons with English, our study proceeded as follows. We first translate all the data set in Warstadt and Bowman (2019) into Korean. Then, we construct three different types of Transformer models (mBERT, KoBERT, and KR-BERT). Then the performances were measured with the basic statistics (precision, recall, F1-score, and accuracy) and the Matthews correlation coefficient (MCC). Through the analysis, it was investigated how differently the various groups of syntactic phenomena behaved depending on the Transformer models.

References

Devlin, J., M. Chang, K. Lee, and K. Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Minneapolis, MN) 1:4171-4186.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and I. Polosukhin. 2017. Attention Is All You Need. arXiv preprint arXiv:1706.03762.

Warstadt, A., A. Singh, and S. Bowma. 2018. Neural Network Acceptability Judgments. arXiv preprint arXiv:1805.12471.

Warstadt, A. and S. Bowman. 2019. Grammatical Analysis of Pretrained Sentence Encoders with Acceptability Judgments. arXiv preprint arXiv:1901.03438.