KoreALBERT: Pretraining a Lite BERT Model for Korean Language Understanding

SAMSUNG SDS

Hyunjae Lee, Jaewoong Yoon, Bonggyu Hwang, Seongho Joe, Seungjai Min, Youngjune Gwon

Milan, Italy 10 | 15 January 2021

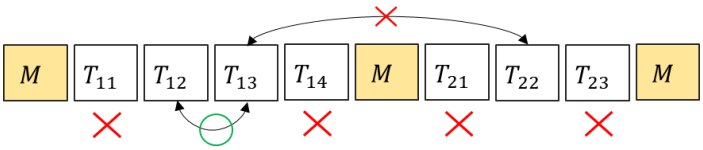
ICPR 2020

Motivation

- Emerging of Pretrained Language Model(PLM) in natural language processing
- Lack of Korean Language Model
- Multi-lingual Pre-trained Model is not sufficient
  → In most cases, Mono-lingual Model outperforms Multi-lingual Model

Contributions

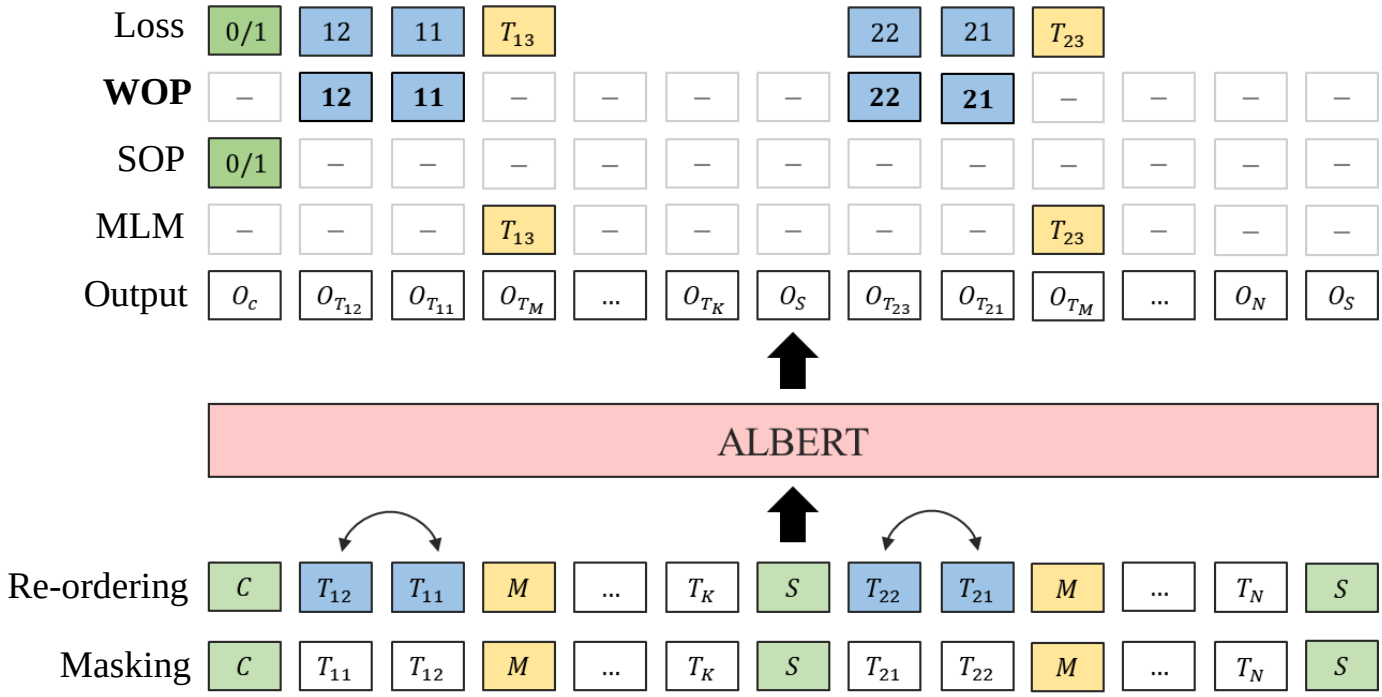
- First Pre-trained Language Model for Korean language using ALBERT Model
- New Training Task : Word Order Prediction



< Token Re-ordering Strategy >

Training Architecture

- Jointly Training → Masked Language Modeling(MLM) + Word Order Prediction(WOP) + Sentence Order Prediction(SOP)



Experimental Results

- KoreALBERT consistently outperforms multi-lingual baselines on 6 downstream NLP tasks while having much fewer parameters.

Model	Params	Speedup	KorNLI	KorSTS	…	NER	KorQuAD	Avg.
Multilingual BERT	172	1.0	76.8	77.8	…	80.3	86.5	83.3
XLM-R	270	0.5x	80.0	79.4	…	**83.9**	92.3	86.4
KoreALBERT	18	1.3x	**81.1**	**82.1**	…	83.7	**94.5**	**87.5**