KoreALBERT: Pretraining a Lite BERT Model for Korean Language Understanding

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

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

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.

<table>
<thead>
<tr>
<th>Model</th>
<th>Params</th>
<th>Speedup</th>
<th>KorNLI</th>
<th>KorSTS</th>
<th>...</th>
<th>NER</th>
<th>KorQuAD</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multilingual BERT</td>
<td>172</td>
<td>1.0</td>
<td>76.8</td>
<td>77.8</td>
<td>...</td>
<td>80.3</td>
<td>86.5</td>
<td>83.3</td>
</tr>
<tr>
<td>XLM-R</td>
<td>270</td>
<td>0.5x</td>
<td>80.0</td>
<td>79.4</td>
<td>...</td>
<td>83.9</td>
<td>92.3</td>
<td>86.4</td>
</tr>
<tr>
<td>KoreALBERT</td>
<td>18</td>
<td>1.3x</td>
<td><strong>81.1</strong></td>
<td><strong>82.1</strong></td>
<td>...</td>
<td>83.7</td>
<td><strong>94.5</strong></td>
<td><strong>87.5</strong></td>
</tr>
</tbody>
</table>