<table>
<tr><th rowspan="2">cases</th><th>doc_1</th><th>doc_2</th><th>decision</th><th>id</th></tr>
<tr>
<td>
<table>
<tr><th>authors</th><td>Nelson F. Liu and Matt Gardner and Yonatan Belinkov and Matthew E. Peters and Noah A. Smith</td></tr>
<tr><th>title</th><td>Linguistic Knowledge and Transferability of Contextual Representations</td></tr>
<tr><th>publication_date</th><td>2019-04-25 00:00:00</td></tr>
<tr><th>source</th><td>SupportedSources.INTERNET_ARCHIVE</td></tr>
<tr><th>journal</th><td></td></tr>
<tr><th>volume</th><td></td></tr>
<tr><th>doi</th><td></td></tr>
<tr><th>urls</th><td>https://web.archive.org/web/20191015035101/https://arxiv.org/pdf/1903.08855v5.pdf</td></tr>
<tr><th>id</th><td>id5183515402127896357</td></tr>
<tr><th>abstract</th><td>Contextual word representations derived from large-scale neural language models are successful across a diverse set of NLP tasks, suggesting that they encode useful and transferable features of language. To shed light on the linguistic knowledge they capture, we study the representations produced by several recent pretrained contextualizers (variants of ELMo, the OpenAI transformer language model, and BERT) with a suite of seventeen diverse probing tasks. We find that linear models trained on top of frozen contextual representations are competitive with state-of-the-art task-specific models in many cases, but fail on tasks requiring fine-grained linguistic knowledge (e.g., conjunct identification). To investigate the transferability of contextual word representations, we quantify differences in the transferability of individual layers within contextualizers, especially between recurrent neural networks (RNNs) and transformers. For instance, higher layers of RNNs are more task-specific, while transformer layers do not exhibit the same monotonic trend. In addition, to better understand what makes contextual word representations transferable, we compare language model pretraining with eleven supervised pretraining tasks. For any given task, pretraining on a closely related task yields better performance than language model pretraining (which is better on average) when the pretraining dataset is fixed. However, language model pretraining on more data gives the best results.</td></tr>
<tr><th>versions</th><td></td></tr>
</table>
</td>
<td>
<table>
<tr><th>authors</th><td>Nelson F. Liu
Matt Gardner
Yonatan Belinkov
Matthew E. Peters
Noah A. Smith</td></tr>
<tr><th>title</th><td>Linguistic Knowledge and Transferability of Contextual Representations</td></tr>
<tr><th>publication_date</th><td>2019-01-01 00:00:00</td></tr>
<tr><th>source</th><td>SupportedSources.INTERNET_ARCHIVE</td></tr>
<tr><th>journal</th><td>Association for Computational Linguistics</td></tr>
<tr><th>volume</th><td></td></tr>
<tr><th>doi</th><td>10.18653/v1/n19-1112</td></tr>
<tr><th>urls</th><td>https://web.archive.org/web/20200505155520/https://www.aclweb.org/anthology/N19-1112.pdf</td></tr>
<tr><th>id</th><td>id7992971970784639212</td></tr>
<tr><th>abstract</th><td>Contextual word representations derived from large-scale neural language models are successful across a diverse set of NLP tasks, suggesting that they encode useful and transferable features of language. To shed light on the linguistic knowledge they capture, we study the representations produced by several recent pretrained contextualizers (variants of ELMo, the OpenAI transformer language model, and BERT) with a suite of sixteen diverse probing tasks. We find that linear models trained on top of frozen contextual representations are competitive with state-of-the-art task-specific models in many cases, but fail on tasks requiring fine-grained linguistic knowledge (e.g., conjunct identification). To investigate the transferability of contextual word representations, we quantify differences in the transferability of individual layers within contextualizers, especially between recurrent neural networks (RNNs) and transformers. For instance, higher layers of RNNs are more taskspecific, while transformer layers do not exhibit the same monotonic trend. In addition, to better understand what makes contextual word representations transferable, we compare language model pretraining with eleven supervised pretraining tasks. For any given task, pretraining on a closely related task yields better performance than language model pretraining (which is better on average) when the pretraining dataset is fixed. However, language model pretraining on more data gives the best results.</td></tr>
<tr><th>versions</th><td></td></tr>
</table>
</td>
<td>DUPLICATES</td>
<td>350</td>
</tr>
</table>