BERT with Pre-train on SQuAD 2.0 Context
Chenchen Pan, Liang Xu
{cpan2, liangxu}@stanford.edu

Introduction
BERT achieves the state-of-the-art results in a variety of language tasks. In this project, we replicated the BERT base model, explored the reason behind the BERT’s strength. We found that the gain comes from the pre-training on large scale corpus, rather than the architecture. However, the pre-training process reduce the model performance on no-answer questions. So we proposed the idea of pre-training on SQuAD 2.0 context to improve this.

Data
SQuAD 2.0 consists of 100k+ question-answer pairs with corresponding passage, and also contains 50k new, unanswerable questions.
Example:
Input Question:
Where do water droplets collide with ice crystals to form precipitation?
Input Paragraph:
... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals within a cloud. ...
Output Answer:
within a cloud

Experiment
No-pretrain vs. Pre-train vs. Pretrain-on-squad

<table>
<thead>
<tr>
<th>Question Type</th>
<th>Number of questions on Dev Set</th>
<th>Number of Exact Match on Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Has Answer</td>
<td>5928</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4270</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4239</td>
</tr>
<tr>
<td>No Answer</td>
<td>5945</td>
<td>5290</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4448</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4277</td>
</tr>
</tbody>
</table>

Approach
To know whether the gain comes from the (1) pre-training process; (2) the use of the large scale unlabeled corpus; (3) the architecture, i.e., self-attention.

Result
The model without pre-training performs better on No-Answer Questions. The model with pre-training performs better on has-answer questions. The strength of BERT model more comes from the pre-training process than the architecture.

Future
Perform the same approach on BERT-large to get to use the full power of the BERT model.
Tune model configuration for currently pre-trained model to achieve better performance.

References