Dealing with long sentences

Abstract

- We participated in the Stance Classification 2 (SC2) subtask of NTCIR-17 QA Lab-PoliInfo-4 as Team KIS.
- We incrementally pretrained the Japanese pretrained LUKE model with a Masked Language Model (MLM) on the Diet minutes dataset.
- We found that these methods were effective, achieved the highest score of 97.41% in accuracy in the formal run of the subtask.

Purpose

- Dealing with long utterances
  - Using head + tail method
- Adaptation to the Japanese political domain
  - Using Incremental pretraining

Approach

1. Dealing with long utterances
   - Using head + tail method
2. Adaptation to the Japanese political domain
   - Using Incremental pretraining

Incremental pretraining

- Politics-specific model
  - Domain-adaptive pretraining (DAPT)
    1. DAPT1 (Minutes-specific model)
       MLM using the full text of the Diet minutes dataset
    2. DAPT2 (Discussion-specific model)
       MLM using the discussion part of the Diet minutes dataset
- Task-specific model
  - Task-adaptive pretraining (TAPT)
    1. TAPT
       MLM using the utterance text portion of the training data

Construction of Japanese politics-specific model

Estimate a politicians’ stance from their utterances.

Dataset

- Stance classification 2 (SC2) dataset
  - The training data
    - 8,534 utterances
  - The test data
    - 2,240 utterances
- Diet minutes dataset
  - Entire set
    - 694,907 blocks
    - Each utterance is divided within 512 tokens considering the period ("。").
  - Discussion part subset
    - 13,204 blocks
    - Utterances containing the Japanese word "discussion (討論)".

Experiments

Dealing with long sentences

- Extracted head (start) and tail (end) of the sentences with different ratios.

<table>
<thead>
<tr>
<th>Head + tail ratio</th>
<th>5-fold Cross Validation</th>
<th>Leader board (Acc)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acc (max)</td>
<td>Acc (min)</td>
</tr>
<tr>
<td>512 / 0</td>
<td>0.9549</td>
<td>0.9349</td>
</tr>
<tr>
<td>384 / 128</td>
<td>0.9654</td>
<td>0.9443</td>
</tr>
<tr>
<td>256 / 256</td>
<td>0.9555</td>
<td>0.9496</td>
</tr>
<tr>
<td>128 / 384</td>
<td>0.9596</td>
<td>0.9420</td>
</tr>
<tr>
<td>0 / 512</td>
<td>0.9653</td>
<td>0.9436</td>
</tr>
</tbody>
</table>

- The combination of 384 tokens and 128 tokens showed a stable performance.

Effectiveness of Domain-Adaptive Pretraining

<table>
<thead>
<tr>
<th>Model</th>
<th>5-fold Cross Validation</th>
<th>Leader board (Acc)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acc (max)</td>
<td>Acc (min)</td>
</tr>
<tr>
<td>without DAPT</td>
<td>0.9654</td>
<td>0.9443</td>
</tr>
<tr>
<td>DAPT1</td>
<td>0.9695</td>
<td>0.9566</td>
</tr>
<tr>
<td>DAPT2</td>
<td>0.9990</td>
<td>0.9490</td>
</tr>
<tr>
<td>DAPT1 + DAPT2</td>
<td>0.9672</td>
<td>0.9537</td>
</tr>
</tbody>
</table>

- DAPT1 showed better performance in both cv and test.
- DAPT1 + DAPT2 showed a best performance in test.

Effectiveness of Task-Adaptive Pretraining

<table>
<thead>
<tr>
<th>Model</th>
<th>5-fold Cross Validation</th>
<th>Leader board (Acc)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acc (max)</td>
<td>Acc (min)</td>
</tr>
<tr>
<td>without DAPT</td>
<td>0.9654</td>
<td>0.9443</td>
</tr>
<tr>
<td>TAPT</td>
<td>0.9883</td>
<td>0.9830</td>
</tr>
<tr>
<td>DAPT1 + TAPT</td>
<td>0.9672</td>
<td>0.9596</td>
</tr>
<tr>
<td>DAPT1 + DAPT2 + TAPT</td>
<td>0.9736</td>
<td>0.9602</td>
</tr>
</tbody>
</table>

- TAPT showed slightly better in test.
- DAPT1 + DAPT2 + TAPT showed a best performance in test.

Conclusion

- We verified the effect of incremental pretraining on a pretrained model with a dataset of the target political domain or task.
- In the future, we would like to apply our incremental pretrained model using the Diet minute dataset to tasks in other political domains to evaluate its generalized performance in the political domain.