CNN for Text-Based Multiple Choice Question Answering

Akshay Chaturvedi¹, Onkar Pandit² and Utpal Garain¹

¹Indian Statistical Institute, Kolkata, India [akshay91.isi@gmail.com, utpal@isical.ac.in] ²INRIA, Lille, France [oapandit@gmail.com]

**Task**

Multiple choice question answering where the question is based on a particular text article.

**Overview**

- The proposed CNN model outperforms several LSTM based baselines on two datasets: TQA and SciQ.
- Question-option tuple as input to generate a score for the concerned option.
- A simple but effective strategy to deal with questions having options like none of the above, two of the above, both (a) and (b) etc.
- Sentence level attention is used instead of word level attention to better capture the important sentences in the article.

**Method**

- The most relevant paragraph is chosen from the text article using the question and options.
- The question option tuple is embedded using CNN consisting of three types of filters of size $f_i \times d \forall j = 1, 2, 3$ with size of output channel as $k$ followed by average pooling.
  
  $$h_i = \text{CNN}(q_i, o_i) \quad \forall i = 1, 2, \ldots, n_q$$

- The sentences in the paragraph are embedded using the same CNN.
  
  $$d_j = \text{CNN}(s_j) \quad \forall j = 1, 2, \ldots, n_{sents}$$

- Using $h_i$, we perform sentence level attention as follows
  
  $$a_{ij} = \frac{h_i \cdot d_j}{||h_i|| \cdot ||d_j||}$$

- To give a score to the $i^{th}$ option, we take the cosine similarity between $h_i$ and $m_i$
  
  $$score_i = \frac{h_i \cdot m_i}{||h_i|| \cdot ||m_i||}$$

- The scores are normalized to get the final probability distribution.
  
  $$p_i = \frac{\exp(score_i)}{\sum_j \exp(score_j)}$$

- We refer to options like none of the above, two of the above, all of the above, both (a) and (b) as forbidden options.

- Let $S = \{score_i \quad \forall i \mid i^{th}$ option not in forbidden options] and $|S| = k$.

1. **Questions with none of the above/ all of the above option:** If $\max(S) - \min(S) < \text{threshold}$ then the final option is the concerned forbidden option.

2. **Questions with two of the above option:** If $S(k) - S(k-1) < \text{threshold}$, then the final option is the concerned forbidden option.

3. **Questions with both (a) and (b) type option:** For these type of questions, let the corresponding scores for the two options be $score_i$ and $score_{i'}$. If $|score_i - score_{i'}| < \text{threshold}$ then the final option is the concerned forbidden option.

4. **Questions with any of the above option:** In this case, we always choose the concerned forbidden option.

- We tried different threshold values ranging from 0 to 1. The threshold was set to that value which gave the highest accuracy on the training set.

**Results**

- Table 1: Accuracy on validation set of TQA dataset.

<table>
<thead>
<tr>
<th>Model</th>
<th>True-False</th>
<th>Multiple Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRU</td>
<td>536/994 (53.9%)</td>
<td>829/1530 (34.6%)</td>
</tr>
<tr>
<td>CNN$_{1,4,5}$</td>
<td>531/994 (52.4%)</td>
<td>531/1530 (34.7%)</td>
</tr>
<tr>
<td>CNN$_{3,4}$</td>
<td>537/994 (54.0%)</td>
<td>543/1530 (35.5%)</td>
</tr>
</tbody>
</table>

- Table 2: Accuracy of the models on SciQ dataset.

<table>
<thead>
<tr>
<th>Model</th>
<th>True-False</th>
<th>Multiple Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>50.9</td>
<td>32.7</td>
</tr>
<tr>
<td>Text-Only</td>
<td>50.2</td>
<td>32.9</td>
</tr>
<tr>
<td>BiDAF</td>
<td>50.4</td>
<td>32.2</td>
</tr>
<tr>
<td>CNN$_{3,4}$</td>
<td>53.7</td>
<td>35.8</td>
</tr>
</tbody>
</table>

- Table 3: Accuracy of different models on TQA dataset.

<table>
<thead>
<tr>
<th>Model</th>
<th>w/o Threshold</th>
<th>Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN$_{3,4}$</td>
<td>109/433</td>
<td>188/433</td>
</tr>
</tbody>
</table>

- Table 4: Threshold strategy on validation set of TQA.

The code is available at https://github.com/akshay107/CNN-QA

**References**
