CNN for Text-Based Multiple Choice Question Answering

Akshay Chaturvedi[1], Onkar Pandit[2] and Utpal Garain[1]

[1]Indian Statistical Institute, Kolkata, India [✉: akshay91.isi@gmail.com, utpal@isical.ac.in]

[2]INRIA, 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 $f_j \times d \; \forall j = 1, 2, 3$ with size of output channel as k followed by average pooling.

$$h_i = CNN([q; o_i]) \quad \forall i = 1, 2, .., n_q$$

- The sentences in the paragraph are embedded using the same CNN.

$$d_j = CNN(s_j) \quad \forall j = 1, 2, .., n_{sents}$$

- Using h_i, we perform sentence level attention as follows

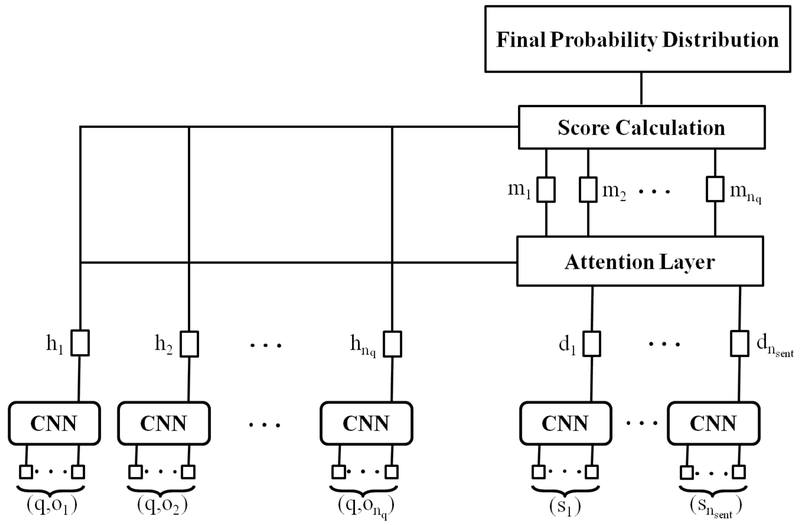

Figure 1: Architecture of our proposed model

$$a_{ij} = \frac{h_i \cdot d_j}{||h_i||.||d_j||}$$

$$r_{ij} = \frac{exp(a_{ij})}{\sum_{j=1}^{n_{sents}} exp(a_{ij})}$$

$$m_i = \sum_{j=1}^{n_{sents}} r_{ij} d_j$$

- To give a score to the i^{th} option, we take the cosine similarity between h_i and m_i

$$score_i = \frac{h_i \cdot m_i}{||h_i||.||m_i||}$$

- The scores are normalized to get the final probability distribution.

$$p_i = \frac{exp(score_i)}{\sum_{i=1}^{n_q} exp(score_i)}$$

- 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 $S = [score_i \; \forall i \mid i^{th}$ option not in forbidden options] and $|S| = k$.
  1. **Questions with none of the above/ all of the above option**: If $max(S) - min(S) < threshold$ then the final option is the concerned forbidden option.
  2. **Questions with two of the above option**: If $S_{(k)} - S_{(k-1)} < threshold$, 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 $score_{i_1}$ and $score_{i_2}$. If $|score_{i_1} - score_{i_2}| < threshold$ 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

Model	True-False	Multiple Choice
GRU_{bl}	536/994 (53.9%)	529/1530 (34.6%)
$CNN_{3,4,5}$	531/994 (52.4%)	531/1530 (34.7%)
$CNN_{2,3,4}$	537/994 (54.0%)	543/1530 (35.5%)

Table 1: Accuracy on validation set of TQA dataset.

Model	Accuracy
GRU_{bl}	68.2%
$CNN_{3,4,5}$	87.1%
$CNN_{2,3,4}$	87.8%
$CNN_{2,3,4}$	84.7% (test-set)

Table 2: Accuracy of the models on SciQ dataset.

Model	True-False	Multiple Choice
Random	50.0	22.7
Text-Only	50.2	32.9
BIDAF	50.4	32.2
$CNN_{2,3,4}$	**53.7**	**35.8**

Table 3: Accuracy of different models on TQA dataset.

Model	w/o Threshold	Threshold
$CNN_{2,3,4}$	109/433	**188/433**

Table 4: Threshold strategy on validation set of TQA.

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

References

[1] Aniruddha Kembhavi, Minjoon Seo, Dustin Schwenk, Jonghyun Choi, Ali Farhadi, and Hannaneh Hajishirzi. Are you smarter than a sixth grader? textbook question answering for multimodal machine comprehension. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[2] Johannes Welbl, Nelson F. Liu, and Matt Gardner. Crowdsourcing multiple choice science questions. In *Proceedings of the 3rd Workshop on Noisy User-generated Text*, pages 94–106, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.