Coattention-Based Multi-Perspective Matching Network for Machine Comprehension

Qiujiang Jin, Bowei Ma

Abstract

Reading comprehension is as the level of understanding of a text/ message. Usually reading comprehension contains the process of reading the context, processing the words or sentences, and then extract conceptual meaning of the whole document. Endowing machines to achieve such capacity is coveted goal for natural language processing. The task of Machine Comprehension (MC) is to enable machine to understand a given paragraph and then answer questions related to the paragraph. In this work, we focus on the Stanford Question Answering dataset (SQuAD) and implemented an end-to-end deep neural network model containing a multi-perspective matching scheme incorporated with coattention to identify the answer span. Experimental result on the test set of SQuAD shows that our model achieves a good training result.

Model Description

Question Answering (QA) Task Definition

- One sample in the SQuAD dataset is represented a triplet (*Q*, *P*, *A*) of question *Q*, paragraph *P* and answer span $A = (a_s, a_e)$.
- Main QA task could be represented by estimating the conditional distribution Pr(*A*| *Q*, *P*) and by independent assumption we may generate answer span by

$$A^* = \underset{1 \le a_s \le a_e \le n}{\arg\max} \Pr(a_s|Q,P)\Pr(a_e|Q,P)$$

Architecture of Multi-Perspective Coattention Network (MPCN)

- **Coattention encoder for (*Q*, *P*)**
  - Run LSTM for (*Q*, *P*) to Generate basic embedding representation (*Q*, *P*).
  - Construct relevancy matrix between (*Q*, *P*) using cosine similarity and weight column *j* of *P* to compute *P'* with weight

$$r_j = \max_{i \in \{1,\cdots,m\}} r_{i,j}$$

  - Compute attention summary (*Q**, *P**) of questions and paragraphs based on the affinity matrix $L = (P')^T Q$.
  - Run a bi-directional LSTM to incorporate contextual information into the representation of each time step.

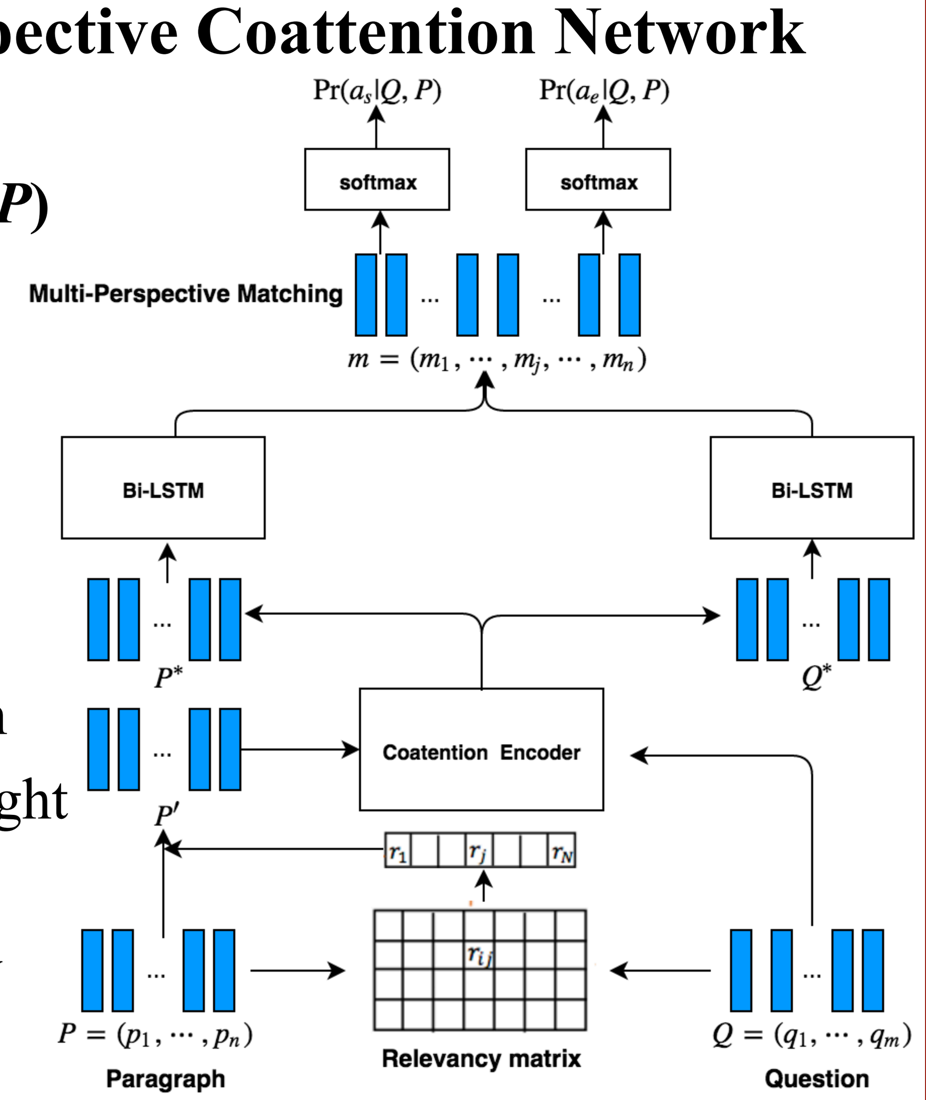

Figure 1: Architecture of MPCN model

- **Multi-Perspective Context Matching**
  - Define multi-perspective embedding comparison function

$$m = f_m(v_1, v_2; W) \in \mathbb{R}^d$$

  - Concatenate the comparison matching of three strategies: Full-Matching, Maxpooling-Matching and Meanpooling-Matching.
  - Aggregate the matching vectors by Bi-LSTM in order to interact one position with its surrounding.
- **Prediction Technique**

  We predict the probability distribution $\Pr(a_s|Q,P), \Pr(a_e|Q,P)$ using feed-forward neural networks and normalize with softmax.

Model Performance

- The model learns well during the first 5 epochs. The loss drops form around 12 to around 3. The F1 and EM score on the train set and validation set rise to 48 and 43 separately.
- But after the epoch 5, the model seems to learn very slowly and those scores seem to not improving a lot even though the model is not overfitting.
- One problem is the small size of the hidden state number and small size of the matching units. The implementation of our model may not reflect the deeper semantic info of the paragraph with respect to the question.

Dataset \ Evaluation Metric	F1 score	Exact-Match Score
Training set (81381 samples)	68.245	55.701
Development set (10000 samples)	49.496	36.272
Test set (9500 samples)	49.878	37.354

Table 1 : Performance of the MPCN model

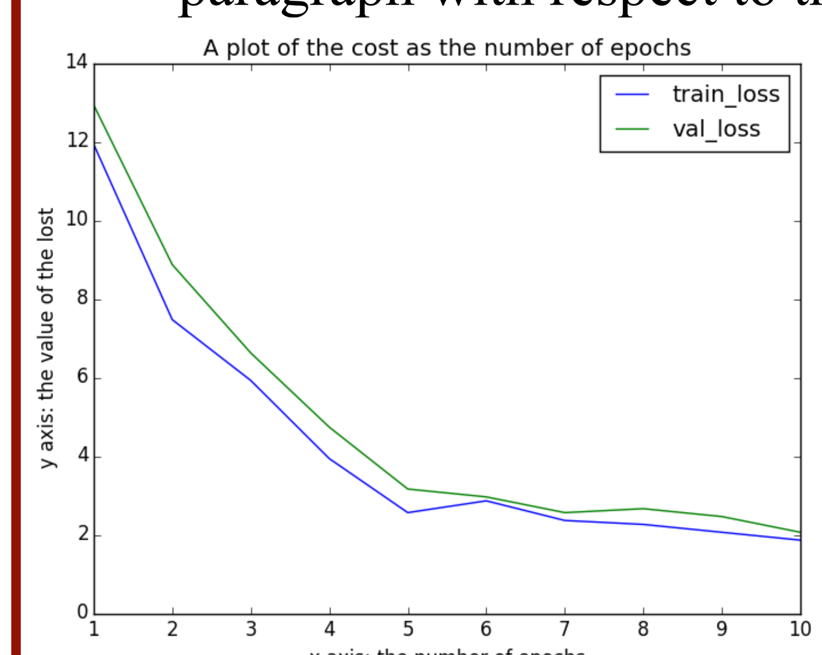

Figure 2: Loss curve during training

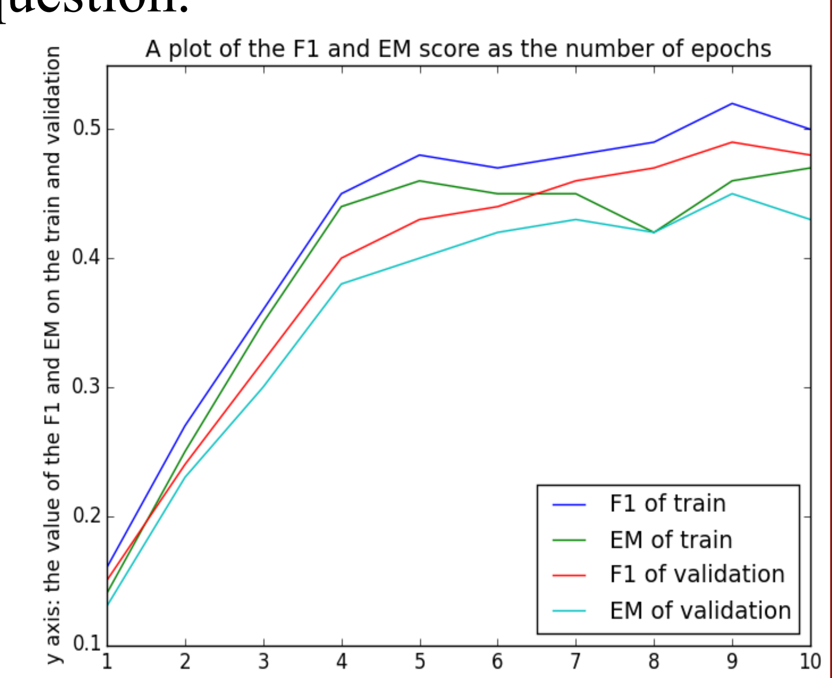

Figure 3: Accuracy curve during training

Error Analysis

- Our model can generally answer questions about time, date ,location and places successfully. For instance,
  - **Question:** What day was the game played on?
  - **Context**: The American Football Conference(AFC) champion Denver Broncos defeated the National Football Conference(NFC) champion Carolina Panthers to earn their third Super Bowl title. The game was played on ***February 7 , 2016*** , at Levi 's Stadium in the San Francisco Bay Area at Santa Clara , California .
- However, our model usually fail to answer other questions where the answer is not adjacent to the exact words of questions in the paragraph. It is also prone to be predicting fairly relevant words to the answer instead of the answer itself.

Discussion

- For many answers, the beginning index is correct but the ending index is wrong. The answers given by this model included more redundant information compared to the correct answer. The decoder layer can be improved to give more accurate estimate of the starting and ending position of the answer.
- Can implement one more layer: the matching of one context hidden state with the average of the question hidden states. This can be a good complement to the existing layer of the average of the matching of one context hidden state with many question hidden states
- Also for the last bidirectional LSTM, may be set the initial states as the representation of the question or context can improve the result.

References

[1] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250*, 2016.

[2] Zhiguo Wang, Haitao Mi, Wael Hamza, and Radu Florian. Multi-perspective context matching for machine comprehension. *arXiv: 1612.04211*, 2016.

[3] Caiming Xiong, Victor Zhong, and Richard Socher. Dynamic coattention networks for question answering. *arXiv:1611.01604*, 2016.