National Laboratory of Pattern Recognition,
Chinese Academy of Sciences

Employing External Rich Knowledge for Machine Comprehension

Bingning Wang, Shangmin Guo, Kang Liu, Shizhu He, Jun Zhao
National Laboratory of Pattern Recognition
Institute of Automation Chinese Academy of Sciences

1 Motivation

Recently proposed machine comprehension (MC) application is an effort to deal with natural language understanding problem. However, the small size of machine comprehension labeled data confines the application of deep neural networks architectures that have shown advantage in semantic inference tasks. Previous methods use a lot of NLP tools to extract linguistic features but only gain little improvement over simple baseline. In this paper, we build an attention-based recurrent neural network model, train it with the help of external knowledge which is semantically relevant to machine compre- hension, and achieves a new state-of-the-art result.

2 The Proposed Method

We denote the document as D and document sentences as $\{s_i, \ldots, s_n\}$, each D has several questions Q = $\{q_0, \ldots, q_i, \ldots, q_m\}$ and each q_i consists of 4 candidate answer $A_i = \{a_{i0}, \ldots, a_{i3}\}$, in order to answer the question, we must choose the relevant sentences S from D and then combine it with the question to get final answer:

$$p(a|q,d) = \underbrace{p(S|q,d)}_{\text{Answer selection}}\underbrace{p(a|q,S)}_{\text{Answer generation}}$$

As we divide the machine comprehension problem as a standard question problem that consists of sub-tasks (i.e. answer selection and answer generation). And these sub-tasks can benefit from off-the-shelf external rich QA resources.

External QA resources:

Answer selection : ***WikiQA, TrecQA, InsuranceQA***

Answer generation (RTE) : ***SICK, SNLI …***

Add external Answer Selection Knowledge

The attention based models has shown great advantage in answer selection tasks, so we adopt a attention based recurrent neural networks architecture to capture the relationship between the question and candidate answers. Particularly, in MCTest, the length of most sentences and questions are no more than 10 tokens, the gradient exploding or vanishing may not be an issue. So we use the simple vanilla type instead of LSTM or GRU as RNN framework:

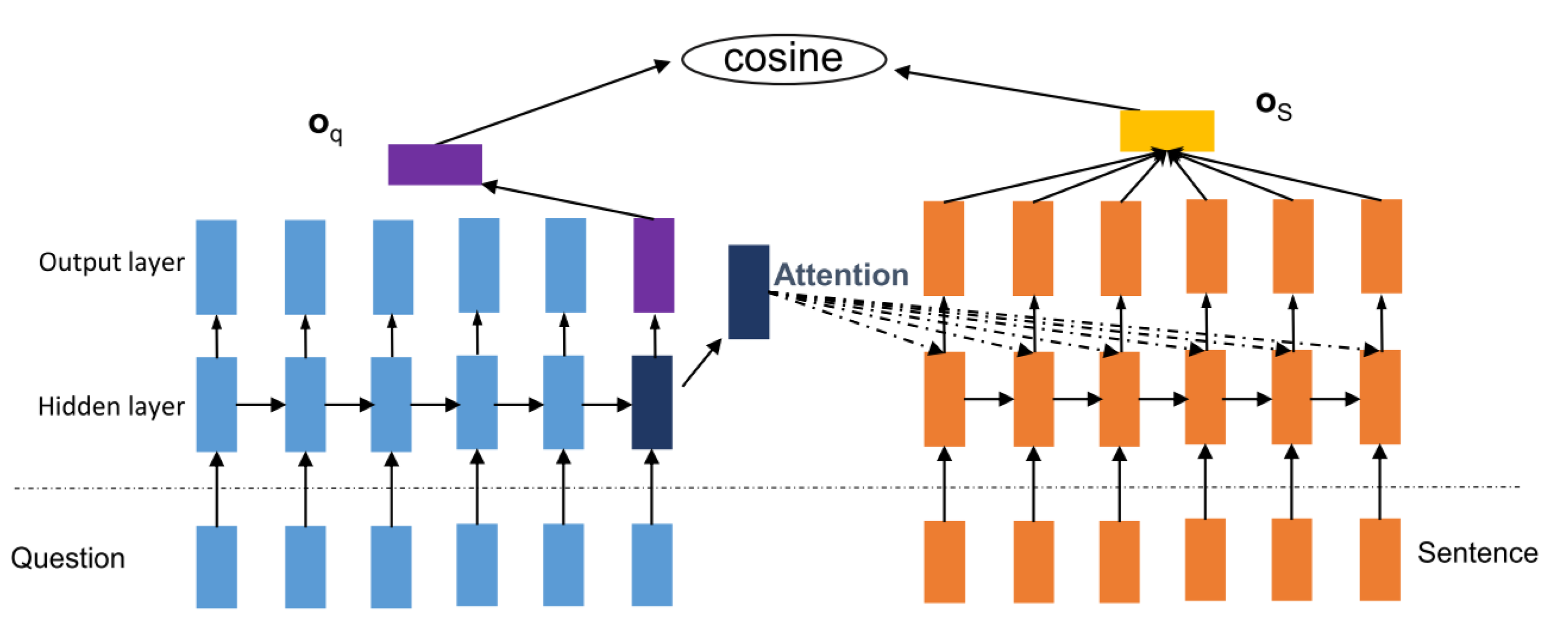


Figure 1: Our answer selection architecture

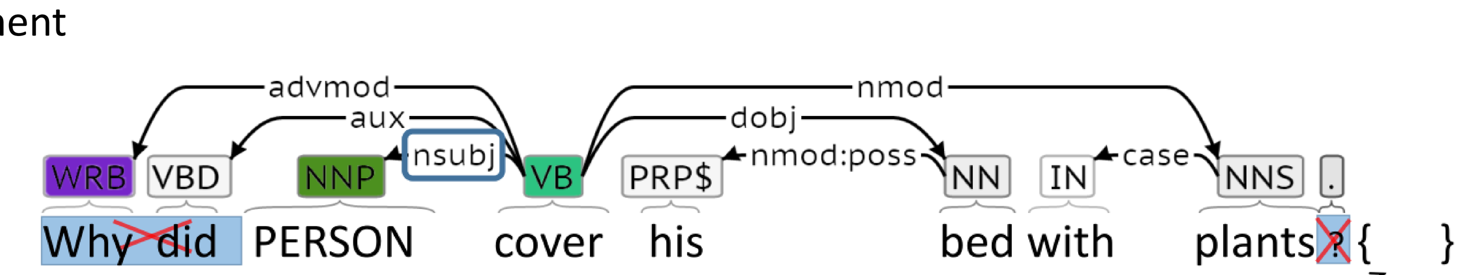
We can employing external rich Answer selection knowledge as an additional supervision to this model:

$$L_{AS}(q,D) = \sum_{s \in D} P(s|q,D;\theta_{RNN}) \log Q(s|q,D)$$

$$L_2(\theta_{+AS}; D_{train}) = \log \sum_{i=1}^{|D_{train}|} \sum_{j=1}^{|Q|} [P(a_{ij}^*|q_{ij}) - \eta L_{AS}(q_{ij}, D_i)] - \lambda g(\theta_{+AS})$$

Add external Answer Generation Knowledge

We first transform each question-answer pair into a statement, and then use an external-RTE-enhanced method to measure the relationship between the sentence and the candidate statement



candidate answer: *{Because he wanted a green bed}*

Figure 2: An example of question transformation

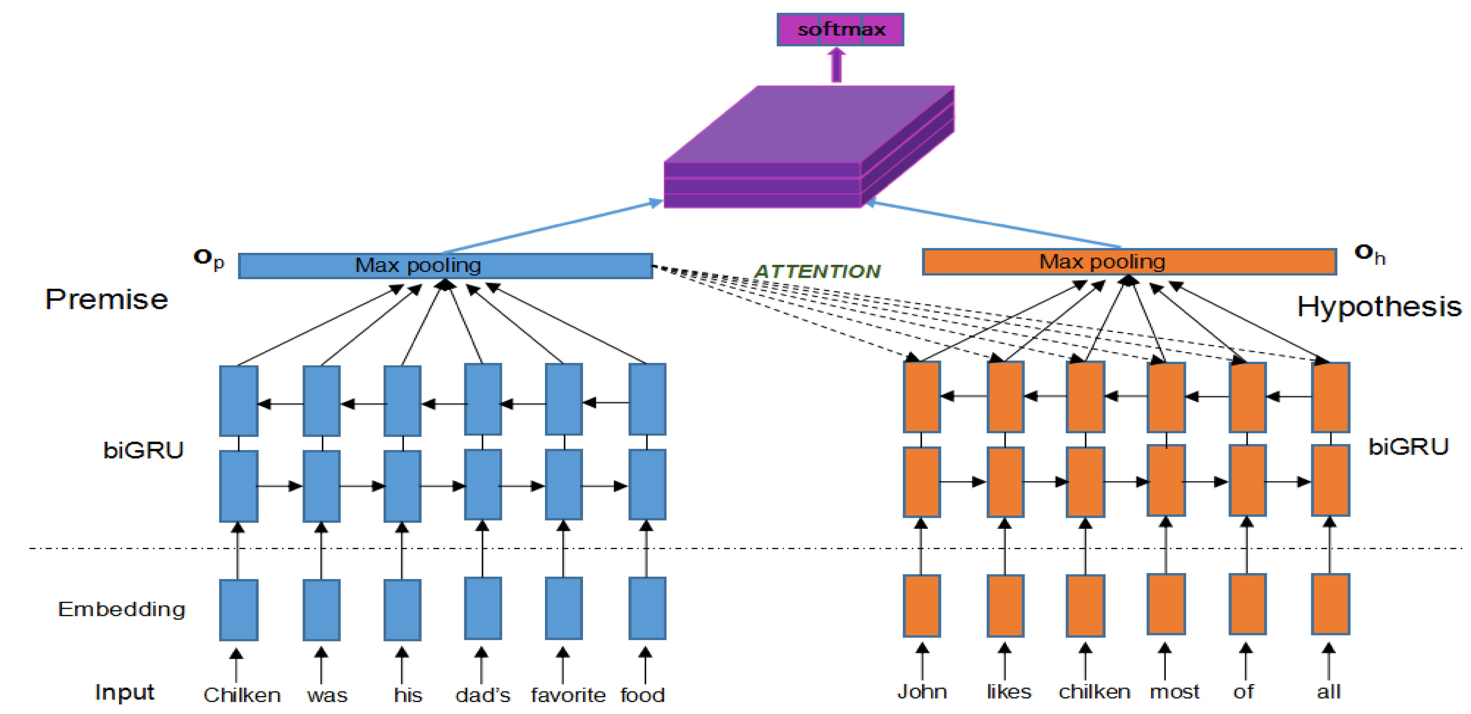


Figure 3: Our RTE architecture

Add Combine with rubust n-gram features

$$P(a|s,D) = [\beta P(s_q|s;\theta_1) + (1-\beta)P(s_q|s;\theta_{RTE})]$$

$\beta = similarity(s_q^-, s)$ Where s_q^- denotes the transformed statement which replaces the answer with a common word ' ANSWER'

$p(s_q|s;\theta_1)$

Constituency match: In constituency tree, subtree are denoted as triplet: a parent node and its two child nodes. We add the number of triplet that **I**: the POS of three nodes are matching. **II**: the head words of parent nodes matching.

Dependency match: In dependency tree, a depen- dency is denoted as (u,v,arc(u,v)) where arc(u,v) de- note dependency relation. We add two terms similarity: **I**:$u_1 = u_2$, $v_1 = v_2$ and arc(u_1, v_1)=arc(u_2, v_2).**II**: whether the root of two dependency tree matches.

Framework

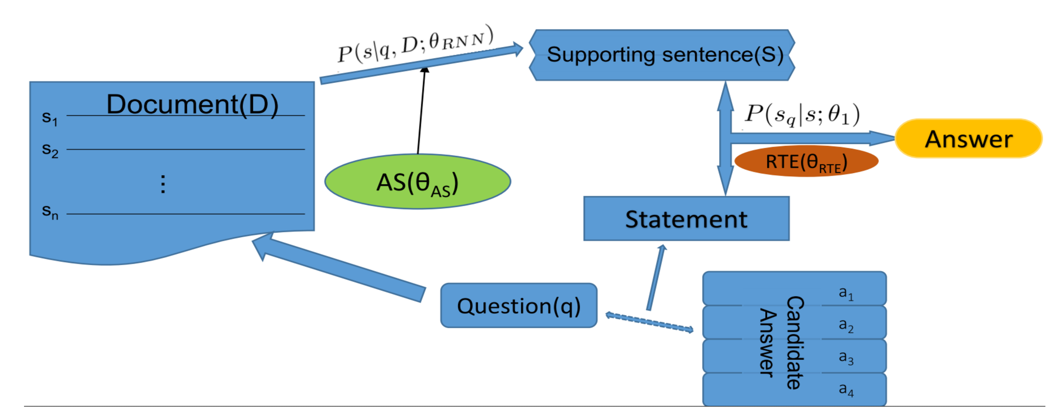


Figure 4: The framework of our approach

3 Experiment

We performed experiments on the MCTest dataset, the result is shown below:

System	MC160			MC500		
	One	Multiple	All	One	Multiple	All
Sliding Window	64.73	56.64	60.41	58.21	56.17	57.09
Sliding Window+Word Distance	75.89	60.15	67.50	64.00	57.46	60.43
Sliding Window+Word Distance+RTE	76.78	62.50	69.16	68.01	59.45	63.33
[Kapashi and Shah, 2015]	-	-	36.0	-	-	34.2
[Narasimhan and Barzilay, 2015]	82.36	65.23	73.23	68.38	59.90	63.75
[Wang and McAllester, 2015]	84.22	67.85	75.27	72.05	67.94	69.94
[Smith *et al.*, 2015]	78.79	**70.31**	75.77	69.12	63.34	65.96
[Sachan *et al.*, 2015]	-	-	-	67.65	**67.99**	67.83
without External Knowledge ($\beta = 1, \eta = 0$)	40.39	37.94	39.08	38.40	33.13	31.33
without External AS knowledge ($\eta = 0$)	41.07	40.63	40.83	49.63	28.05	32.83
without External RTE knowledge ($\beta = 1$)	74.11	64.06	68.75	57.72	50.91	53.00
Final Model	**88.39**	64.84	**75.83**	**79.04**	63.51	**70.96**

Table 1: Results on MC500 and MC160

The external RTE model and answer selection model result on SNLI and WikiQA compared with state of the art are shown in Table 2.

	Answer Selection		RTE
	MAP	MRR	Accuracy
State of the Art	0.6921	**0.7108**	**0.835**
Our method	**0.6936**	0.7094	0.829

Table 2: Results on external resource

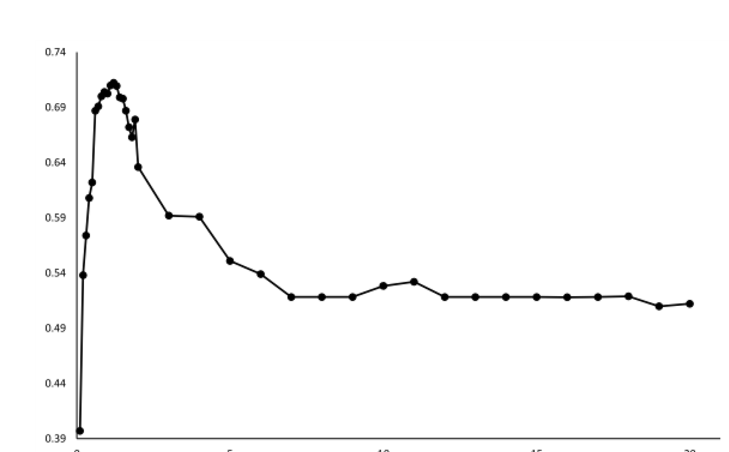


Figure 5: The result of different η Value in MC500

We also check the impact of η value which is shown in Figure 5

4 Conclusions

In this paper, for the subpart of MC process, we build an attention-based RNN model for AS process and add external RTE model to answer generation process. To build a deep learning model from limited data, we train the model with supervision from external knowledge, customize the external resources and add it to MC process properly. The experiment result shows that our model achieves especially well in single support fact question. Error analysis suggests that modeling the relationship between sentences in AS can yield improvement on this task. In addition, the counting problem and common sense problem are really hard to tackle which requires deeper linguistic analysis. In the future, we plan to build a RTE model that could model multiple sentences together for inference tasks.