Reasoning with Memory Augmented Neural Networks for Language Comprehension, ICLR17

Tsendsuren Munkhdalai & Hong Yu

UMass Medical School

ICLR 2017
Presenter: Jack Lanchantin

Outline

Hypothesis Testing

- In QA, a hypothesis is formed by regressing the original query, and then the hypothesis is tested against reality (support set)
- If the model is satisfied with the current hypothesis, the reasoning process is stopped and the answer is found. Else, reformulate hypothesis until the answer is found.

Cloze-Type QA

Context

The BBC producer allegedly struck by Jeremy Clarkson will not press charges against the "Top Gear" host, his lawyer said Friday. Clarkson, who hosted one of the most-watched television shows in the world, was dropped by the BBC Wednesday after an internal investigation by the British broadcaster found he had subjected producer Oisin Tymon "to an unprovoked physical and verbal attack." ...

Query

Producer **X** will not press charges against Jeremy Clarkson, his lawyer says.

Answer

Oisin Tymon

Previous Work

- Many similar models for this task: Attentive Reader (Hermann et. al. 2015, Memory Networks (Sukhbaatar et. al. 2015), EpiReader (Trischler et al 2016)
- Previous models designed with the predefined number of computational hops for inference, but different types of questions require different types of reasoning

Outline

Memory Initialization

$$\text{Query Memory: } M_0^q = BiLSTM^q(Q) \in \mathbb{R}^{k \times |Q|} \quad (1)$$

$$\text{Document Memory: } M^d = BiLSTM^d(D) \in \mathbb{R}^{k \times |D|} \quad (2)$$

Outline

Updating Memory through Hypothesis-Test Loop

- Hypothesis Formulation: In each step, the query memory M^q is updated with content from the document memory to form a new query
- Hypothesis Testing: The new query is then checked against the document facts to make an answer prediction
- 3 Steps:
  1. Read
  2. Compose
  3. Write

1. Read

Given Query state $s_t^q \in \mathbb{R}^k$, and Document state $s_t^d \in \mathbb{R}^k$:

$$\text{Read vector } (k): \quad r_t = LSTM^r([s_{t-1}^q; s_{t-1}^d]) \tag{3}$$

$$\text{Q Alignment } (|Q|): \quad l_t^q = r_t^T M_{t-1}^q \tag{4}$$

$$\text{New Q State } (k): \quad s_t^q = softmax(l_t^q)^T M_{t-1}^q \tag{5}$$

$$\text{Q Memory Key } (|Q|): \quad z_t^q = sigmoid(l_t^q) \tag{6}$$

$$\text{D Alignment } (|D|): \quad l_t^d = s_t^q T M^d \tag{7}$$

$$\text{New D State } (k): \quad s_t^d = softmax(l_t^d)^T M^d \tag{8}$$

2. Compose

Combines the current query state s_t^q, document state s_t^d, and hidden state of read module r_t, resulting in single vector c_t

$$c_t \in \mathbb{R}^k = MLP^c(s_t^q, s_t^d, r_t) \tag{9}$$

3. Write

Takes outputs z_t^q (Q Memory Key), and s_t^d (Document State) from the read module and updates the query memory:

$$M_t^q = M_{t-1}^q z_t^q + s_t^d (1 - z_t^q) \tag{10}$$

Outline

Ending The Hypothesis-Test Loop

- Need to decide when to stop the read-compose-write loop and answer
- 2 strategies are explored:
  1. Query-Gating
  2. Adaptive Computation

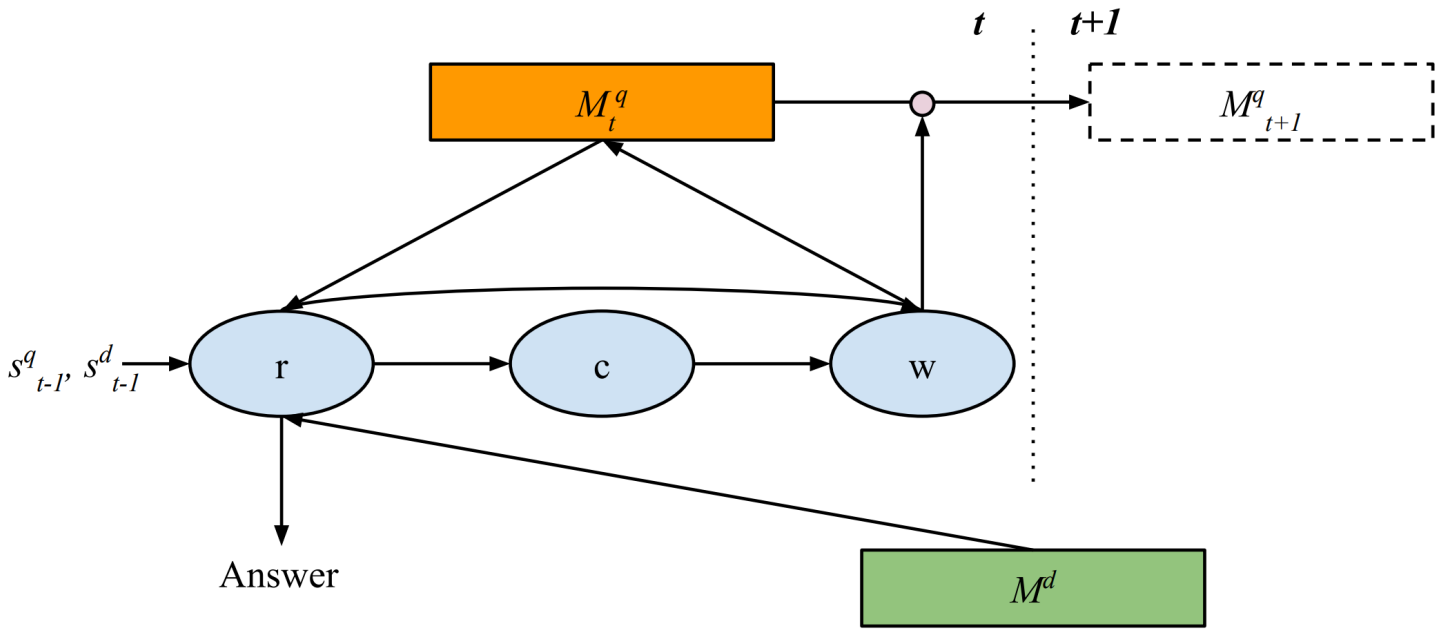
Query Gating

Instead of making a hard decision on halting the loop (i.e. stop reading), the write module performs a word-level query gating as:

$$\text{Write vector } (k): \quad w_t = LSTM^w(c_t) \tag{11}$$

$$\text{Write gate } (k): \quad g_t^q = sigmoid(w_t^T M_{t-1}^q) \tag{12}$$

$$\text{New Query Memory } (k \times |Q|): \quad M_t^q = M_t^q(1 - g_t^q) + M_{t-1}^q g_t^q \tag{13}$$



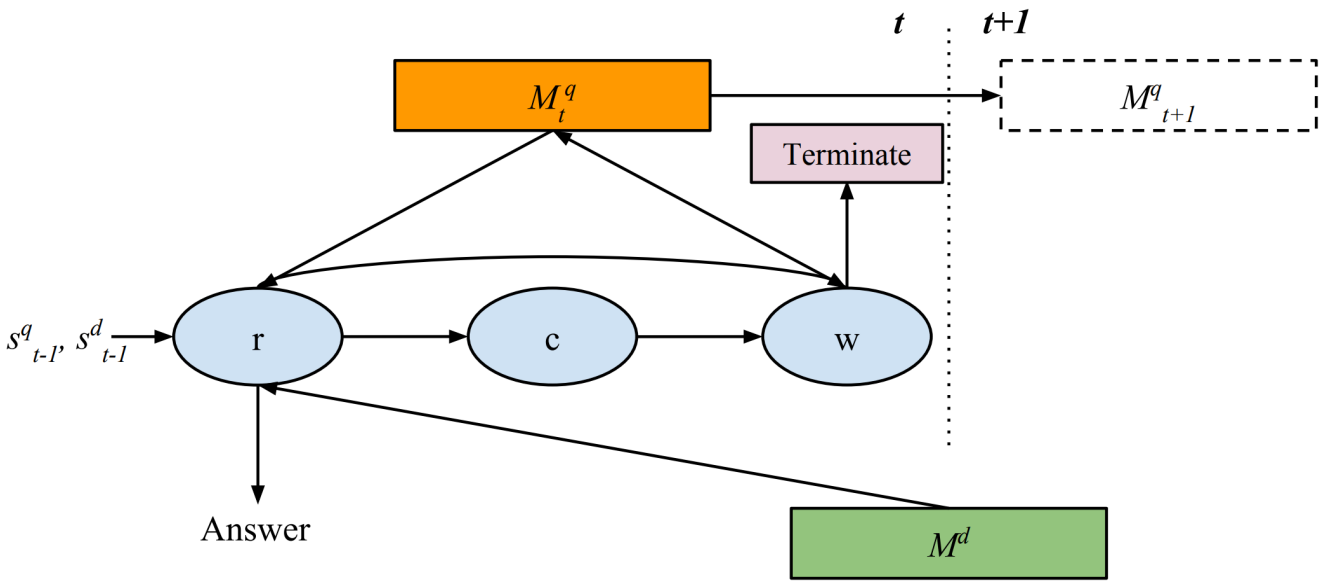
Adaptive Computation

Write module is equipped with a termination, which decides its willingness to continue or finish in each step:

$$\text{Write vector } (k): \quad w_t = LSTM^w(c_t) \tag{14}$$

$$\text{Termination Score (scalar)}: \quad e_t = sigmoid(o^T w_t) \tag{15}$$

$$\text{End probability}: \quad p_t = e_t \prod_{i=1}^{t-1}(1 - e_i), \quad p_T = 1 - \sum_{i=1}^{T-1} p_i \tag{16}$$



Answer Prediction

$$P_t(a|Q, D) = v^T softmax(I_t^d) \tag{17}$$

where $v \in \mathbb{R}^{|D|}$ is a mask denoting the positions of the answer token a in the document (ones for the token and zeros otherwise). I_t^d is the query-document alignment score

Children's Book Test (CBT)

S: 1 Mr. Cropper was opposed to our hiring you .
2 Not , of course , that he had any personal objection to you , but he is set against female teachers , and when a Cropper is set there is nothing on earth can change him .
3 He says female teachers ca n't keep order .
4 He 's started in with a spite at you on general principles , and the boys know it .
5 They know he 'll back them up in secret , no matter what they do , just to prove his opinions .
6 Cropper is sly and slippery , and it is hard to corner him . ''
7 `` Are the boys big ? ''
8 queried Esther anxiously .
9 `` Yes .
10 Thirteen and fourteen and big for their age .
11 You ca n't whip 'em -- that is the trouble .
12 A man might , but they 'd twist you around their fingers .
13 You 'll have your hands full , I 'm afraid .
14 But maybe they 'll behave all right after all . ''
15 Mr. Baxter privately had no hope that they would , but Esther hoped for the best.
16 She could not believe that Mr. Cropper would carry his prejudices into a personal application .
17 This conviction was strengthened when he overtook her walking from school the next day and drove her home .
18 He was a big , handsome man with a very suave , polite manner .
19 He asked interestedly about her school and her work , hoped she was getting on well , and said he had two young rascals of his own to send soon .
20 Esther felt relieved .

q: She thought that Mr. _____ had exaggerated matters a little .

C: Baxter, Cropper, Esther, course, fingers, manner, objection, opinion, right, spite.

a: Baxter

- 20 consecutive sentences from a childrens story
- 21st sentence in which a word has been deleted
- list of ten choices for the deleted word
- tests story completion rather than reading comprehension

Childrens Book Test (CBT) Results

Table 2: Model comparison on the CBT dataset.

	CBT-NE		CBT-CN	
Model	dev	test	dev	test
Human (context + query) (Hill et al., 2015)	-	81.6	-	81.6
LSTMs (context + query) (Hill et al., 2015)	51.2	41.8	62.6	56.0
MemNNs (window mem. + self-sup.) (Hill et al., 2015)	70.4	66.6	64.2	63.0
AS Reader (Kadlec et al., 2016)	73.8	68.6	68.8	63.4
GA Reader (Dhingra et al., 2016)	74.9	69.0	69.0	63.9
EpiReader (Trischler et al., 2016)	75.3	69.7	71.5	67.4
IAA Reader (Sordoni et al., 2016)	75.2	68.6	72.1	69.2
AoA Reader (Cui et al., 2016)	77.8	72.0	72.2	69.4
MemNN (window mem. + self-sup. + ensemble) (Hill et al., 2015)	70.4	66.6	64.2	63.0
AS Reader (ensemble) (Kadlec et al., 2016)	74.5	70.6	71.1	68.9
EpiReader (ensemble) (Trischler et al., 2016)	76.6	71.8	73.6	70.6
IAA Reader (ensemble) (Sordoni et al., 2016)	76.9	72.0	74.1	71.0
NSE ($T = 1$)	76.2	71.1	72.8	69.7
NSE Query Gating ($T = 2$)	76.6	71.5	72.3	70.7
NSE Query Gating ($T = 6$)	77.0	71.4	73.0	**72.0**
NSE Query Gating ($T = 9$)	78.0	72.6	73.5	71.2
NSE Query Gating ($T = 12$)	77.7	72.2	**74.3**	71.9
NSE Adaptive Computation ($T = 2$)	77.1	72.1	72.8	71.2
NSE Adaptive Computation ($T = 12$)	**78.2**	**73.2**	74.2	71.4

CBT-NE: Named Entities, CBT-CN: Common Nouns

Who-Did-What (WDW)

Passage: Tottenham won 2-0 at Hapoel Tel Aviv in UEFA Cup action on Thursday night in a defensiv
impressed Spurs skipper Robbie Keane. ... Keane scored the first goal at the Bloomfield Stadium with D
who insisted earlier on Thursday he was happy at the London club, heading a second. The 26-year-old Berba
reports linking him with a move had affected his performances ... Spurs manager Juande Ramos has won th
the last two seasons ...

Question: Tottenham manager Juande Ramos has hinted he will allow XXX to leave if the Bulgaria striker
is unhappy.

Choices: (1) Robbie Keane (2) Dimitar Berbatov

Who-Did-What (WDW) Results

Table 3: Model comparison on the WDW dataset.

	Strict		Relaxed	
Model	dev	test	dev	test
Human (Onishi et al., 2016)	-	84.0	-	-
Attentive Reader (Hermann et al., 2015)	-	53.0	-	55.0
AS Reader (Kadlec et al., 2016)	-	57.0	-	59.0
GA Reader (Dhingra et al., 2016)	-	57.0	-	60.0
Stanford Attentive Reader (Chen et al., 2016)	-	64.0	-	65.0
NSE ($T = 1$)	65.1	65.5	66.4	65.3
NSE Query Gating ($T = 2$)	65.4	65.1	65.7	65.5
NSE Query Gating ($T = 6$)	65.5	65.7	65.6	65.8
NSE Query Gating ($T = 9$)	65.8	65.8	65.8	65.9
NSE Query Gating ($T = 12$)	65.2	65.5	65.7	65.4
NSE Adaptive Computation ($T = 2$)	65.3	65.4	66.2	66.0
NSE Adaptive Computation ($T = 12$)	**66.5**	**66.2**	**67.0**	**66.7**

Conclusions

- Hypothesis Testing QA
- Query memory gating which prevents from forgetting old query and adaptive computation with termination head