Paper Number: 2444

SMU SINGAPORE MANAGEMENT UNIVERSITY
School of Computing and Information Systems

Context Modeling with Evidence Filter for Multiple-Choice Question Answering

Sicheng Yu, Hao Zhang, Jing Wei, Jing Jiang

Motivation

Motivation

- Most of methods for multiple-choice question answering typically encode each option with the context independently.
- The ratio of evidence sentences is quite low.

Observation

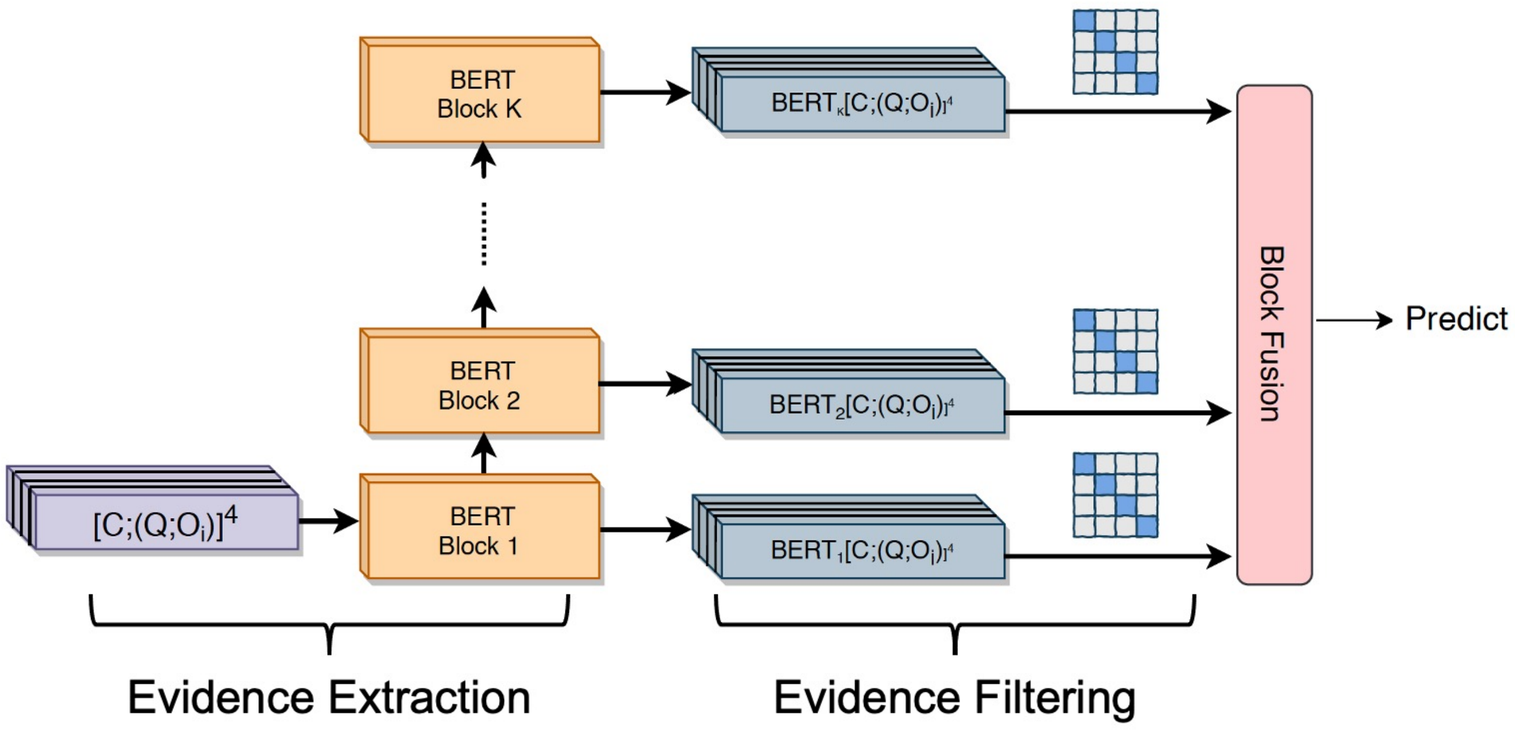
- If a sentence in the context has a similar level of relevance on all of the given options, then it is highly likely that this sentence is not useful for answering the question.
- An evidence sentence in the context is likely to be closely related to the correct option but irrelevant to the incorrect options.

Question: What is the best way to guess a baby's eye color?

A. The surroundings they are born in.
B. Their parents' usual diet.
C. Just take a random guess.
D. The genealogy records of their family.

Context: It is an academic guess too. Hypothesis means scientific guess about the cause and effect of an event. **Eye color is an inherited characteristic.** Some monkey babies can be raised with two parents. Ancestors are family members. The color of an object is the result of the way the object reflects or emits light. **Having offspring produces a family.** Adults have babies. Seals are born on waterfronts. Sugars are important for a plant's diet. Climate is the usual kind of weather in a location. Babies need milk to live. Frog babies in sacs are in eggs. Animals take in oxygen. The crust is just above the mantle. The vision organ is the eye. A person's diet determines nutrient levels. An omnivore includes animals in its diet. Drought is a period of less than usual precipitation.

Solution Architecture



- Evidence Extraction: implicitly extract evidence from context.
- Evidence Filtering: adjust the evidence by considering the relationship between evidence with respect to options.

Experiment Results

Methods	OpenbookQA Dev (%)	OpenbookQA Test (%)
Question Match + ELMo [1]	54.6	50.2
Odd-one-out Solver [1]	56.9	50.2
ESIM + ELMo [1]	53.9	48.9
OFT [12]	-	52.0
OFT (ensemble) [12]	-	52.8
Reading Strategies+GPT [12]	-	55.2
Reading Strategies+GPT (ensemble) [12]	-	55.8
BERT-large (leaderboard)	-	60.4
BERT(large) Multi-task (leaderboard)	-	63.8
Ours Model	**66.8**	**65.6**

Main Result

Modification	Accuracy (%)
(1) w/o block fusion; w/o evidence filter	60.0
(2) w/o block fusion; evidence filter w/o constraints	63.8
(3) w/o block fusion; evidence filter	65.0
(4) block fusion with same evidence filter	64.0
block fusion with different evidence filter (ours)	65.6

Ablation Study

Index	1	2	3	4	5	6	7	8	9	10	11	12
α	1.3418	1.3418	1.3418	1.3418	1.3418	1.3418	1.3408	1.3389	1.3447	1.3457	1.3447	1.3389
β	-1.0693	-1.0693	-1.0693	-1.0693	-1.0693	-1.0693	-1.0703	-1.0723	-1.0664	-1.0664	-1.0664	-1.0723
Index	13	14	15	16	17	18	19	20	21	22	23	24
α	1.3379	1.3457	1.3369	1.3477	1.3467	1.3457	1.3398	1.3467	1.3477	1.3477	1.3486	1.3408
β	-1.0723	-1.0654	-1.0742	-1.0635	-1.0645	-1.0654	-1.0713	-1.0654	-1.0635	-1.0635	-1.0625	-1.0713

Evidence Filter Parameter

Conclusions

- We propose evidence filter to alleviate the effect of unrelated sentences and enhance the saliency of evidences potentially without human efforts.
- Results on OpenbookQA indicate the effectiveness of our method.