

# Context Modeling with Evidence Filter for Multiple-Choice Question Answering

*Sicheng Yu<sup>1</sup>, Hao Zhang<sup>2</sup>, Wei Jing<sup>3</sup>, Jing Jiang<sup>1</sup>*

<sup>1</sup>Singapore Management University, Singapore

<sup>2</sup>Nanyang Technological University, Singapore

<sup>3</sup>Alibaba Group, China

# Motivation

## Task Definition of MCQA

Pick the best matched answer given the context, question and options.

## 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.

*How can we utilize the evidence sentences effectively with the help of 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.

# Motivation

## 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.

*A method to capture the differences among context sentences with respect to the options.*

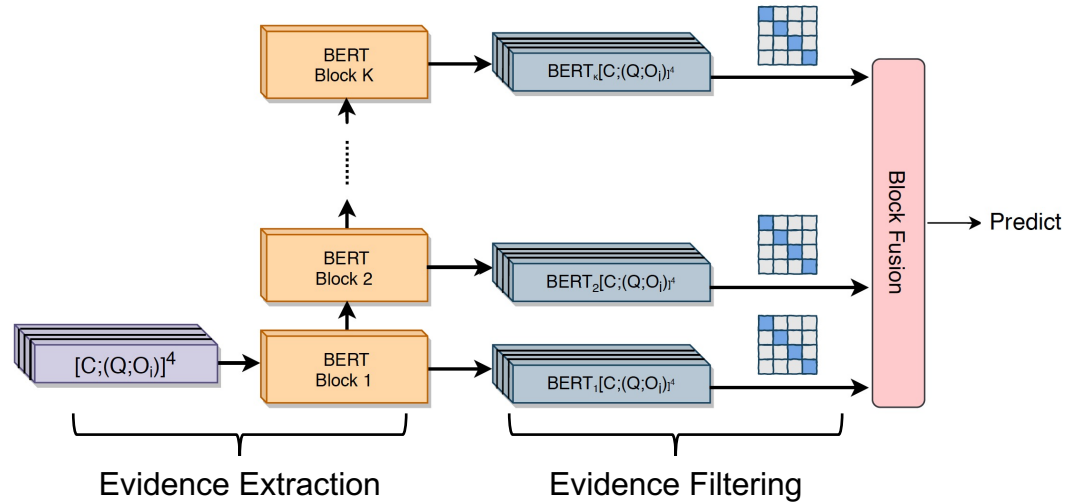
**Question:** What is the best way to guess a baby's eye color?

- 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.

# Method

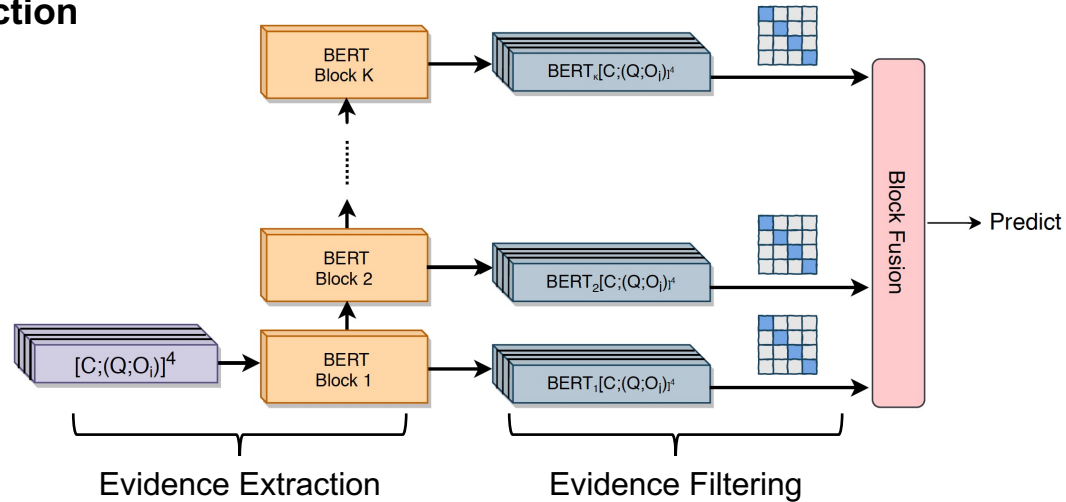
## Overall



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

# Method

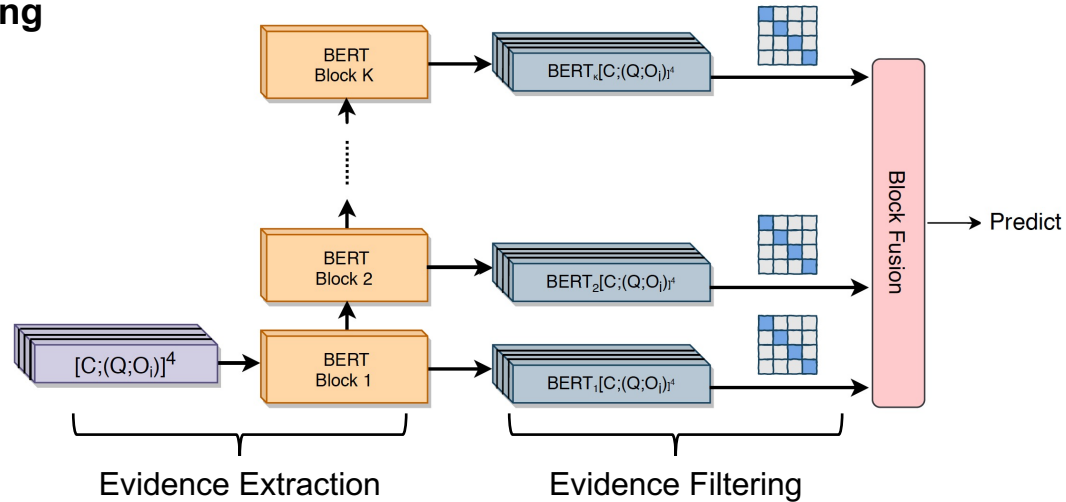
## Evidence Extraction



- Input: context  $C$ , question  $Q$  with one of the option, here we use four options as example.
- BERT: contextual feature and evidence extractor with  $K$  blocks, here we use BERT-large where  $K=24$ .

# Method

## Evidence Filtering



- After obtaining the representation from BERT, we fed them into evidence filter matrix.
- Evidence filter matrix ( $4 \times 4$ ): model the context and filter the evidence with respect to four options:

$$\mathbf{A} \cdot BERT_k([C; (Q; O_i)]^4)$$

# Method

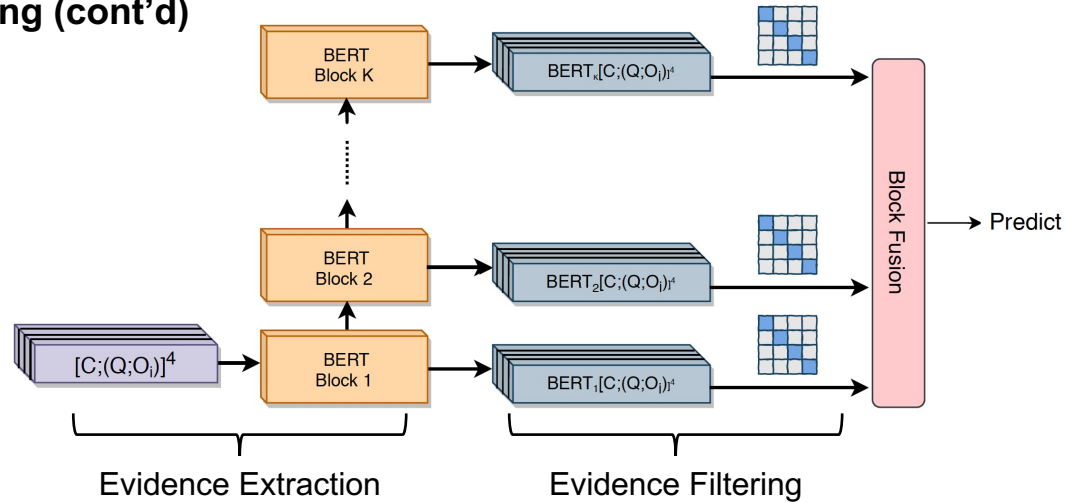
## Evidence Filter Matrix



- If randomly and individually initializing each entry, shuffling options may get different answer.
- To address it, we apply the constrain on the evidence filter matrix.

# Method

## Evidence Filtering (cont'd)



- Block fusion layer: integrate the intermediate output from each evidence filter matrix by a single linear layer.
- Prediction: linear layer for multiple-choice task.



# Experiments

## Main results

- The BERT-based model also shows stable superiority over the models based on other pre-trained language model such as ELMo and GPT.
- Despite the simplicity and few additional parameters of our model, it still outperforms other BERT based approaches.

Methods	OpenbookQA	
	Dev (%)	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
<b>Ours Model</b>	<b>66.8</b>	<b>65.6</b>

# Experiments

## Ablation study

- It demonstrates the effectiveness of evidence filter by comparing (1) with (2)-(4).
- The results from (3), (4) and the last one (ours) suggest that block fusion sharing the same evidence filter performs worse than the model without block fusion, while performs better when different evidence filters are applied.

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

# Conclusion

- 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.

Thank You!