Extraction of Question-related Sentences for Reading Comprehension Tests via Attention Mechanism

Junjie Shan, Yoko Nishihara, Ryosuke Yamanishi, Akira Maeda

2020-12-03

RITSUMEIKAN

Contents

Background

- Question Generation (QG) task tends to use neural network-based Seq2Seq method in recent years.
- Seq2Seq method needs large amount of Sentence-Question pairs (S-Q pairs) in training complex neural network models.
- Existing methods chose to use the SQuAD dataset with manually marked-up S-Q pairs as the training data
  - SQuAD was provided for QA task, sometimes not fit for the reading comprehension of language tests
  - manually prepare S-Q pairs is a labor/time-costing work
- Extraction of question-related sentences from given reading materials is a meaningful & important work for QG task

QA & Reading Comprehension Question

- Questions in QA-based SQuAD dataset (could find answers directly)

Related Sentence(s)	Question
There are 11 cities in the country.	how many cities are there in the country?

- Questions in Reading Comprehension-based RACE dataset (often need to obtain answers by understanding)

Related Sentence(s)	Question
From Chengdu to Jiuzhaigou, you can travel along the East Line not only the other line going past Dujiangyan and Wenchuan. It only takes about three hours to Jiangyou. ... Pingwu has the best royal Buddhist temples in Ming Dynasty, Bao'en Temple and the historic relics in the period of Three Kingdoms. Here we would like to recommend another hotel in Jiuzhaigou for you, ...	how many cities are there in the passage?

Contributions

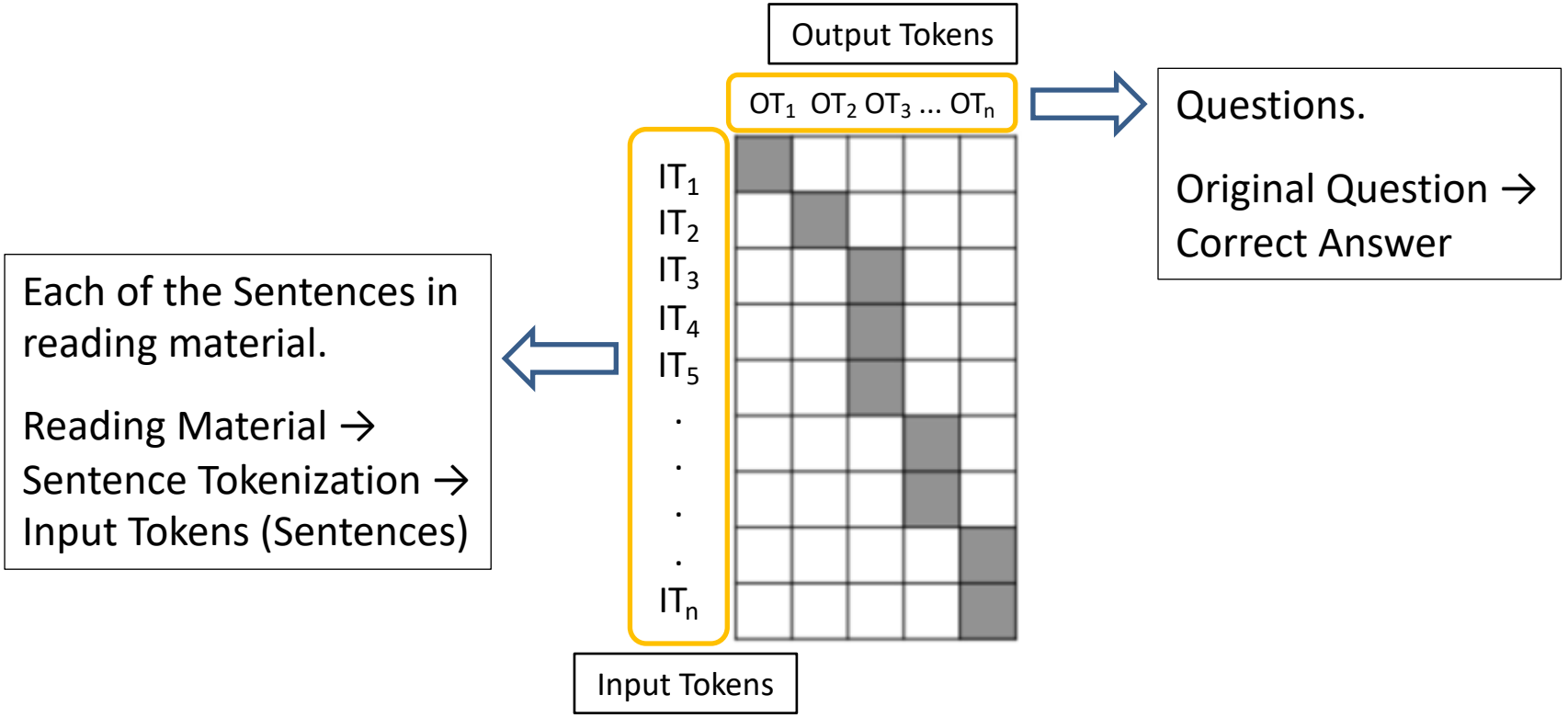
- Use attention mechanism to extract question-related sentences from given reading materials
  - Attention model input: sentences in given reading material
  - Attention model output: reading comprehension questions

- Tested two types of embedding level at output part and evaluated the extracted results respectively
  - Output of Sentence-Level embedding (S2S)
  - Output of Word-Level embedding (S2W)

- Compared extraction results with using traditional Cosine similarity method to extract question-related sentences

Proposed Method

- Attention Mechanism & Attention Scores
  - Attention mechanism uses variable weights (attention scores) to encode input tokens
  - Every output token has a list of attention scores shows the relevant degrees with all of the input tokens
  - Attention scores could be trained through the training data



Input Part & Output Part

- Input Reading Material → Sentence Tokens

Input Article:
I know that for many of you, today is the first day of school. Some of you are probably wishing it were still summer and you could have stayed in bed just a little bit longer this morning. But I'm here because I want to talk with you about your education and what's expected of all of you in this new school year. Every single one of you has something that you're good at. …

↓

S1: [I know that for many of you, today is the first day of school.],
S2: [Some of you are probably wishing it were still summer and you could have stayed in bed just a little bit longer this morning.],
S3: [But I'm here because I want to talk with you about your education and what's expected of all of you in this new school year.],
S4: [Every single one of you has something that you're good at.]
…

- Original Question → Correct Answer

Original Question	Options	Correct Answer
The bus school will _ .	A. take children to school B. stay there for lunch C. take the fathers and mothers to school D. go round from place to place (Correct: D)	The bus school will go round from place to place.
Which of the following is true of the introduction of tea into Britain?	A. The Britons got expensive tea from India. B. Tea reached Britain from Holland. C. The Britons were the first people in Europe who drank tea. D. It was not until the 17th century that the Britons had tea. (Correct: B)	Tea reached Britain from Holland.

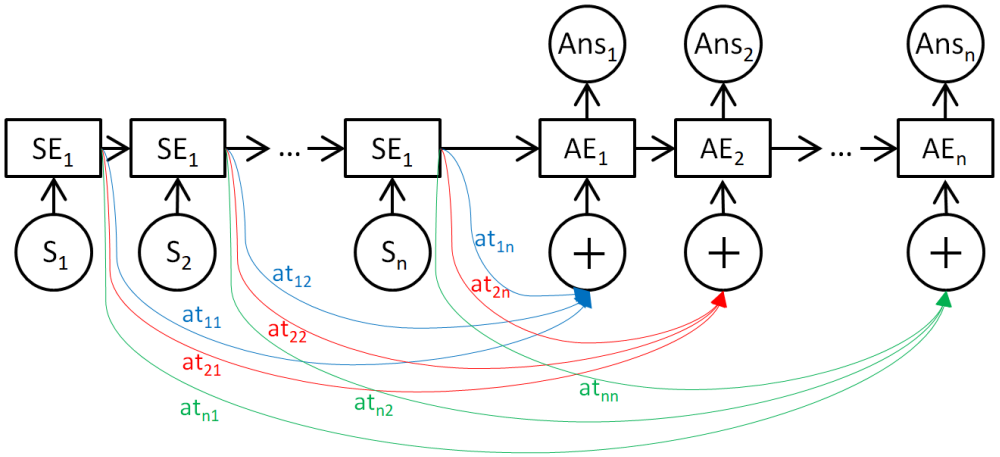
S2S Attention

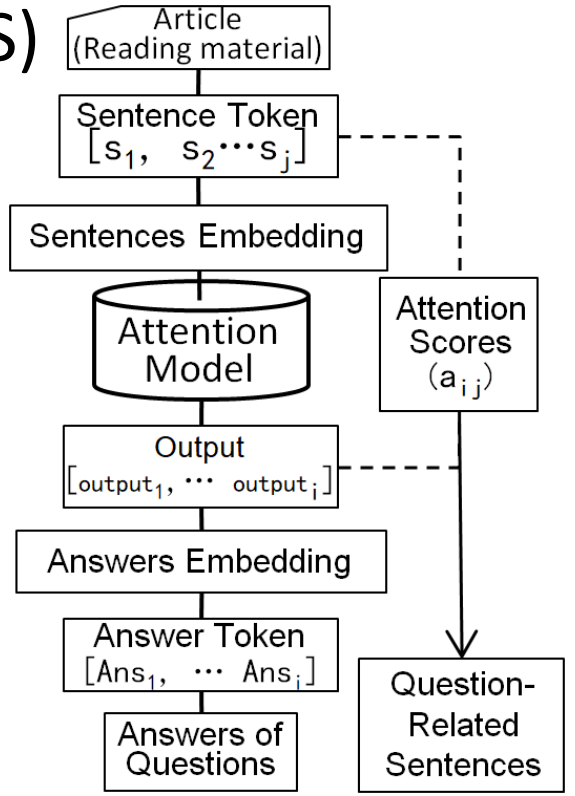
- Output token is one answer-represented question
  - sentence-level to sentence-level (S2S)
    - Each output token (answer) have a list of attention scores relates to all input tokens.

$$at_{ij} = Softmax(Attention(S_j, Output_i))$$

$$Attention(S_j, Output_i) = Min(Cos(Ans_i, Output_i))$$

$$R(s_j|Ans_i) = at_{ij}$$





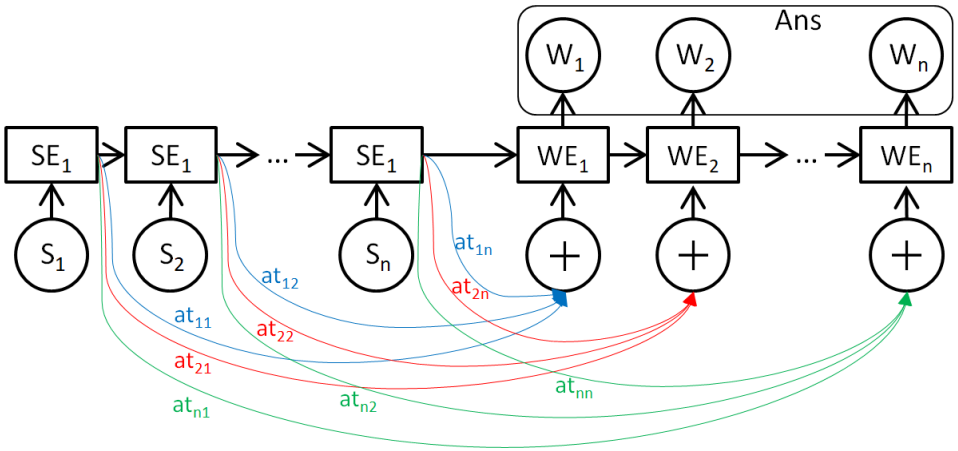
a_{ij}: attention score of $output_i \rightarrow S_j$
S: Sentences in reading material
Ans: question-represented Answer
SE: Input Sentence Embedding
AE: Answer Embedding

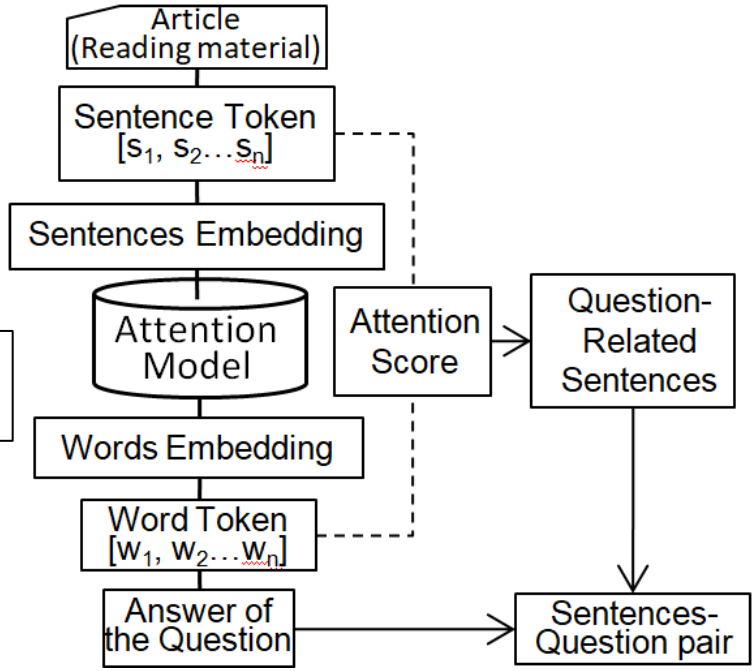
S2W Attention

- Output token is one word of the answer
  - sentence-level to word-level (S2W)
    - Each output token (word) have a list of attention score relates to all input tokens.

$$R(s_j|Ans) = \sum_{i=0,\ldots,n} at_{ij}.$$

a_{ij}: attention score of output $Word_i \rightarrow S_j$





Comparison Method (Cosine Similarity)

- Cosine Similarity of each correct answer and the sentences of input reading material
  - Attention Model was trained to minimize the Cosine Distance between outputs and correct answers.
  - Embedded by same BERT pre-trained model.

$$R(s_j | Ans_i) = COS(SE_j, AE_i)$$

$$SE_j = \mathrm{BERT}(s_j)$$

$$AE_i = \mathrm{BERT}(Ans_i)$$

S: Sentences in reading material
Ans: question-represented Answer
SE: Input Sentence Embedding
AE: Answer Embedding

Extraction Sample of Question-related Sentences

Given Article:

My husband is a born shopper. He loves to look at things and to touch them. He likes to compare prices between the same items in different shops. He would never think of buying anything without looking around in several different shops. On the other hand, I'm not a shopper. I think shopping is boring and unpleasant. If I like something and I have enough money to take it, I buy it at once. I never look around for a good price or a better deal. Of course my husband and I never go shopping together. Doing shopping together would be too painful for both of us. When it comes to shopping, we go our different ways. Sometimes I ask my son Jimmy to buy some food in the shop not far from our home. But he is

saw me he said, "I'm sorry, Mum. I have forgotten to buy oranges and the meat. I only remembered to buy six eggs, but I've dropped three of them." (Totally 31 sentences.)

Question:

They never go shopping together because _ .

Correct Answer:

They never go shopping together because their ways of shopping are quite different.

Underline Sentences: (Related, #4)
Manually marked-up question-related sentences.

Totally 31 Sentences: (Sum, #31)
Number of sentences in the given article.

Word-Level: (Hit 1)
He loves to look at things and to touch them.
Doing shopping together would be too painful for both of us.
When it comes to shopping, we go our different ways.

Sentence-Level: (Hit 3)
He likes to compare prices between the same items in different shops.
When it comes to shopping, we go our different ways.
Of course my husband and I never go shopping together.

Cosine Similarity: (Hit 2)
He likes to compare prices between the same items in different shops.
Of course my husband and I never go shopping together.
But he is always absent-minded.

Extract each of the top three related sentences for one question.
→ In manually pre-checked 41 questions, the average amount of the related sentences is 2.2 .

	TP	FP	Related	Sum
s2w	1	2	4	31
s2s	3	0	4	31
cosine similarity	2	1	4	31

TP=Hit: correctly extracted sentences
FP=3-TP: extract top three for each
Related: all question-related sentences
Sum: Total sentences number

Extraction Results

- Extraction results from same 50 questions.

	TP	FP	Related	Sum	Precision	Recall	Accuracy	F1
s2s	**69**	81	103	991	0.46	0.669	**0.883**	0.54
s2w	**32**	118	103	991	0.213	0.31	**0.809**	0.25
cosine similarity	**45**	105	103	991	0.3	0.437	**0.835**	0.35

  - s2s > cos > s2w
  - s2w: sentence-level to word-level made a deviation
  - s2s & cos: same BERT pre-trained model for embedding
    - A speculation: the attention model did a fine-tuning-like step in the training process. BERT pre-trained model was trained on a more generic dataset.

Conclusion & Future Work

- Extract question-related sentences from given reading materials through attention mechanism.
- Embedding input tokens & output tokens of the attention model in same sentence-level could yield the highest accuracy of extraction.
- Similar methods may also be used in extracting the content related to the given target for other fields.
- Future works:
  - Try improve the extraction accuracy with other index
  - Extract question-related sentences with flexible strategy instead of a fixed number of three

Thank you!