

Generating Distractors for Reading Comprehension Questions from Real Examinations

Yifan Gao¹, Lidong Bing², Piji Li³, Irwin King¹, Michael R. Lyu¹

¹Department of Computer Science and Engineering, The Chinese University of Hong Kong

²R&D Center Singapore, Machine Intelligence Technology, Alibaba DAMO Academy ³Tencent AI Lab

¹{yfgao, king, lyu}@cse.cuhk.edu.hk ²l.bing@alibaba-inc.com ³pijili@tencent.com



code & data

What is 'Distractor Generation'?

A New Task: Generate reasonable distractors (wrong options) for multiple choices questions (MCQs) in reading comprehension

Definition: Given an article P containing multiple sentences s_1, s_2, \dots, s_n , a pair of question q and its correct option a originated from the article, the goal is to generate the distractor d :

$$\bar{d} = \arg \max_d \log P(d|P, a, q)$$

Example:

Article:

The Yanomami live along the rivers of the rainforest in the north of Brazil. They have lived in the rainforest for about 10,000 years and they use more than 2,000 different plants for food and for medicine. But in 1988, someone found gold in their forest, and suddenly 45,000 people came to the forest and began looking for gold. **They cut down the forest to make roads. They made more than a hundred airports.** The Yanomami people lost land and food. **Many died because new diseases came to the forest with the strangers.**

In 1987, they closed fifteen roads for eight months. No one cut down any trees during that time. In Panama, the Kuna people saved their forest. **They made a forest park which tourists pay to visit.** The Gavioes people of Brazil use the forest, but they protect it as well. They find and sell the Brazil nuts which grow on the forest trees.

Question:

Those people built roads and airports in order to _____.

- A. carry away the gold conveniently (**Answer**)
- B. make people there live a better life (**Distractor**)
- C. stop spreading the new diseases (**Distractor**)
- D. develop the tourism there (**Distractor**)

Real-world Applications:

- Help the preparation of **MCQ reading comprehension datasets**
 - The existence of distractors fail existing content-matching SOTA reading comprehension on MCQs like RACE dataset (Lai et al. 2017)
 - Large datasets can **boost the performance** of MCQ reading comprehension systems
- Alleviate instructors' workload in **designing MCQs** for students
 - Poor distractor options can make the questions almost trivial to solve
 - Reasonable distractors are time-consuming to design

Difference with Previous Works

Previous:

- Extract a fixed distractor candidate set (usually word phrases)
- Use similarity-based methods or learning-based methods to select the distractors (a classification or ranking problem)

Our Goal:

- Generate **longer and semantic-rich** distractors (8.1 tokens on avg.)
- The generated distractors should **semantically related** to the reading comprehension question
- The distractors should **not be paraphrases** of the correct answer option
- The generated distractors should be **grammatically consistent with the question**, especially for questions with a blank in the end

Dataset

RACE Reading Comprehension Dataset:

- A large-scale ReAding Comprehension dataset from Examinations (RACE) that are created for middle school and high school students

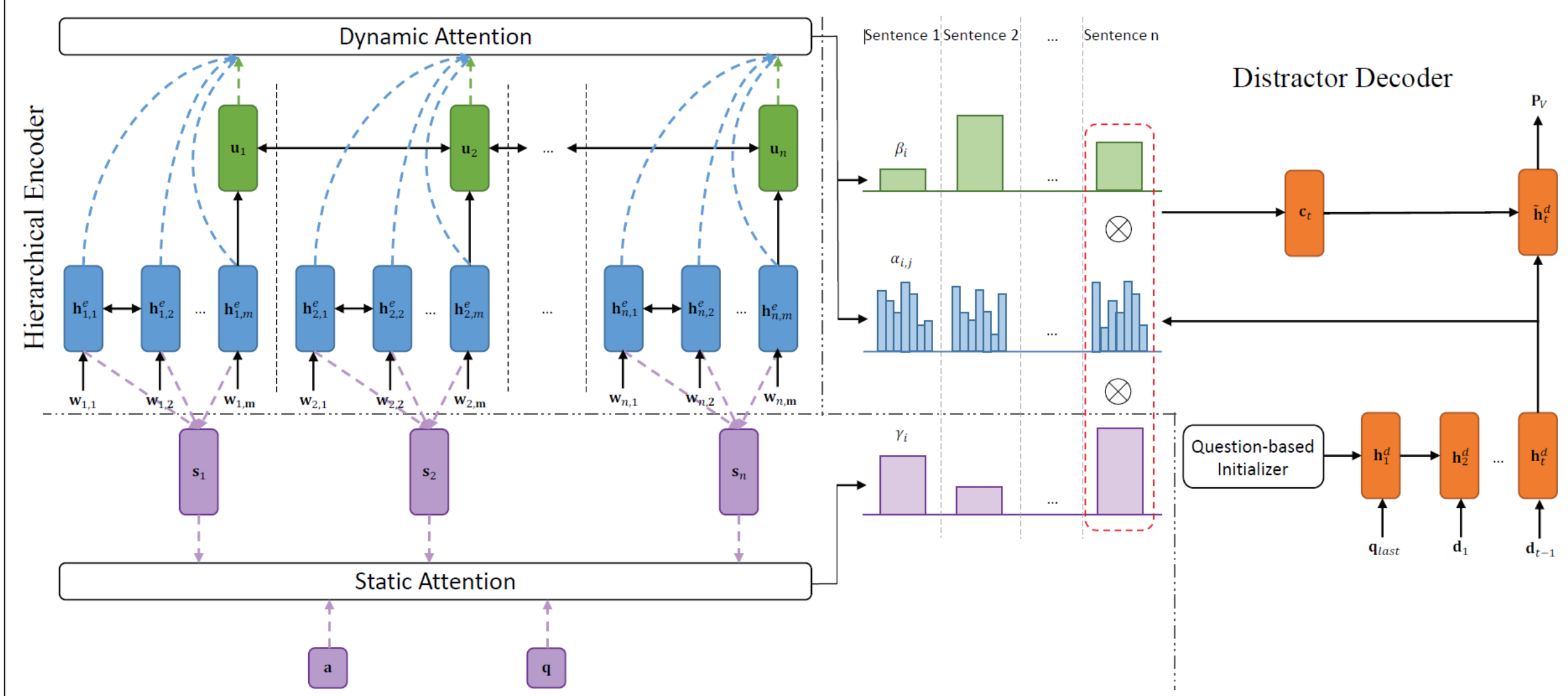
Pruning for our Distractor Generation Task:

- Some distractors have **no semantic relevance** with the article, which can be easily excluded
- Some distractors require **some world knowledge** outside the article
- Pruning Constraint: Keep a distractor if its **weighted frequency of meaningful tokens** is larger than 5

Our Dataset Statistics:

# Train Samples	96501
# Dev Samples	12089
# Test Samples	12284
Avg. article length (tokens)	347.0
Avg. distractor length	8.5
Avg. question length	9.9
Avg. answer length	8.7
Avg. # distractors per question	2.1

Model



1. Static Attention: learn an importance distribution $(\gamma_1, \gamma_2, \dots, \gamma_n)$ of the sentences (s_1, s_2, \dots, s_n) in the article

- Encoding: $\mathbf{a} = \frac{1}{k} \sum_{t=1}^k \mathbf{a}_t, \mathbf{q} = \frac{1}{l} \sum_{t=1}^l \mathbf{q}_t, \mathbf{s}_i = \frac{1}{m} \sum_{t=1}^m \mathbf{h}_{i,t}^e$
- Matching: $o_i = \lambda_q \mathbf{s}_i^T \mathbf{W}_m \mathbf{q} - \lambda_a \mathbf{s}_i^T \mathbf{W}_m \mathbf{a} + \mathbf{b}_m$
- Normalization: $\tau = \text{sigmoid}(\mathbf{w}_q^T \mathbf{q} + b_q), \gamma_i = \text{softmax}(\frac{o_i}{\tau})$

2. Dynamic Attention: $\alpha_{i,j} = \mathbf{h}_{i,j}^e \mathbf{W}_{d_2} \mathbf{h}_t^d, \beta_i = \mathbf{u}_i^T \mathbf{W}_{d_1} \mathbf{h}_t^d$

3. Modulation: use the static attention to modulate the dynamic hierarchical attention $\tilde{\alpha}_{i,j} = \frac{\alpha_{i,j} \beta_i \gamma_i}{\sum_{ij} \alpha_{i,j} \beta_i \gamma_i}$

4. Question-based Initializer:

- Use the last token in the question as the initial input of the decoder
- Use the final cell state and hidden state of the question LSTM to initialize the decoder

5. Generate Diverse Distractors: Use beam search to receive k candidate distractors, select top 3 with predefined Jaccard distance

Experiment Results

Automatic Evaluation:

		BLEU ₁	BLEU ₂	BLEU ₃	BLEU ₄	ROUGE ₁	ROUGE ₂	ROUGE _L
1st Distractor	Seq2Seq	25.28	12.43	7.12	4.51	14.12	3.35	13.58
	HRED	26.10	13.96	8.83	6.21	14.83	4.07	14.30
	Our Model	27.32	14.69	9.29	6.47	15.69	4.42	15.10
2nd Distractor	Seq2Seq	25.13	12.02	6.56	3.93	13.72	3.09	13.20
	HRED	25.18	12.21	6.94	4.40	13.94	3.11	13.40
	Our Model	26.56	13.14	7.58	4.85	14.72	3.52	14.15
3rd Distractor	Seq2Seq	25.34	11.53	5.94	3.33	13.78	2.82	13.23
	HRED	25.06	11.69	6.26	3.71	13.65	2.84	13.04
	Our Model	26.92	12.88	7.12	4.32	14.97	3.41	14.36
Avg. Performance	Seq2Seq	25.25	11.99	6.54	3.92	13.87	3.09	13.34
	HRED	25.45	12.62	7.34	4.77	14.14	3.34	13.58
	Our Model	26.93	13.57	8.00	5.21	15.13	3.78	14.54

Manual Evaluation:

	Annotator 1	Annotator 2	Annotator 3	# Selected
Seq2Seq	31	35	30	96
HRED	33	40	35	108
Our Model	43	45	36	124
Human	75	70	79	224

Case Study

Article:

1. Dear friends, The recent success of children's books has made the general public aware that there's a huge market out there.
2. And there's a growing need for new writers trained to create the \$3 billion worth of children's books bought each year... plus stories and articles needed by over 650 publishers of magazines for children and teenagers.
3. Who are these needed writers?
4. They're ordinary people like you and me.
5. But am I good enough?
6. I was once where you might be now.
7. My thoughts of writing had been pushed down by self-doubt, and I didn't know where to turn for help.
8. Then, I accepted a free offer from the Institute to test my writing ability, and it turned out to be the inspiration I needed.
9. The promise that paid off The Institute made the same promise to me that they will make to you, if you show basic writing ability: you will complete at least one manuscript suitable to hand in to a publisher by the time you finish our course.
10. I really didn't expect any publication before I finished the course, but that happened.
11. I sold three stories.
12. And I soon discovered that was not unusual at the Institute.
13. Since graduation, I have written 34 nationally published children's books and over 300 stories and articles.
14. Free test and brochure We offer a free ability test and will send you a copy of our brochure describing our recognized home-study courses on the basis of one-on-one training.
15. Realize your writing dream today.
16. There's nothing sadder than a dream delayed until it disappears forever.
17. Sincerely, Kristi Hill Institute of Children's Literature

Question 1: You are promised to publish one manuscript when you _____.
Options:
A. show basic ability B. finish the course
C. have sold three stories D. have passed the test
Seq2Seq:
1. have made a mistake
2. have written a lot of books
3. have been writing a newspaper
HRED:
1. have finished the course 1. have sold three stories
2. have a free test 2. write a book
3. have been opened 3. have passed the test
Our Model:
1. To introduce the home-study courses.
2. To show she has realized her dream.
3. To prove she is a qualified writer.
Seq2Seq:
1. To show she is a successful publisher.
2. To show how inspiring her books are.
3. To show her interest in writing books.
HRED:
1. To encourage readers to buy more books.
2. To show she wanted to improve her reading skills.
3. To prove she is a well-known courses publisher.
Our Model:
1. To prove she is a qualified writer.
2. To show her great achievements in literature.
3. To encourage readers to be interested in writing.

Static Attention Distribution

Reference

1. Lai, Guokun et al. "RACE: Large-scale ReAding Comprehension Dataset From Examinations." EMNLP (2017).