



Improving Question Generation with to the Point Context

Jingjing Li¹, Yifan Gao¹, Lidong Bing², Irwin King¹ and Michael R. Lyu¹

¹The Chinese University of Hong Kong, Shatin, N.T. Hong Kong

²R&D Center Singapore, Machine Intelligence Technology, Alibaba DAMO Academy

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



Introduction

Task Definition:

- Question generation (QG) is the task of generating a question from a reference sentence and a specified answer within the sentence. The generated question should be relevant to the given answer.

Existing Works:

- Use BIO tagging scheme to enrich word representation of answer phrases
- Proximity based answer position encoding

Limitation of Existing Works:

- Surrounding words of answer phrase are not necessarily useful context in the complex natural languages.

Sentence: The daily mean temperature in January, the area's coldest month, is 32.6 °F (0.3 °C); however, temperatures usually drop to 10 °F (-12 °C) several times per winter and reach 50 F (10 °C) several days each winter month.

Gold Question: What is New York City's daily January mean temperature in degrees celsius ?

Proximity-based Model: What is the coldest temperature in Celsius ?

Structured Answer-relevant Relation: (The daily mean temperature in January; is; 32.6 °F (0.3 °C))

- The proximity between a answer phrase and context words can not effectively indicate the semantic interaction between them.

Distance	B1	B2	B3	B4	MET	R-L
0~10 (72.8% of #)	45.25	30.31	22.06	16.54	21.54	46.26
>10 (27.2% of #)	35.67	21.72	14.82	10.46	16.72	37.63

(Distance: relative distance between the answer fragment and other non-stop sentence words that also appear in the ground truth question.)

Our Contribution

We propose a question generation system to combine unstructured sentences and structured answer-relevant relations for generation

- The proposed framework can be applied as an extension of other question generation model
- Given multiple facts within one sentence, our model can generate diverse questions by verifying the input of relation encoder

Problem Formulation

S is the reference sentence;

A is a contiguous text span within the S ;

M is the one of the answer-relevant facts within sentence S .

Given the sentence S , the answer A and the answer-relevant relation M , the question system aims to generate the best question \bar{Q} such that:

$$\bar{Q} = \underset{Q}{\operatorname{argmax}} \operatorname{Prob}(Q|S, A, M)$$

Answer-Relevant Relation Extraction

Sentence: The daily mean temperature in January, the area's coldest month, is 32.6 °F (0.3 °C); however, temperatures usually drop to 10 °F (-12 °C) several times per winter and reach 50 F (10 °C) several days each winter month.

Structured Answer-relevant Relation:

- 0.95 (The daily mean temperature in January; is; 32.6 °F (0.3 °C))
- 0.94 (temperatures; drop; to 10 °F (12 °C); several times per winter; usually)
- 0.90 (temperatures; reach; 50 °F)

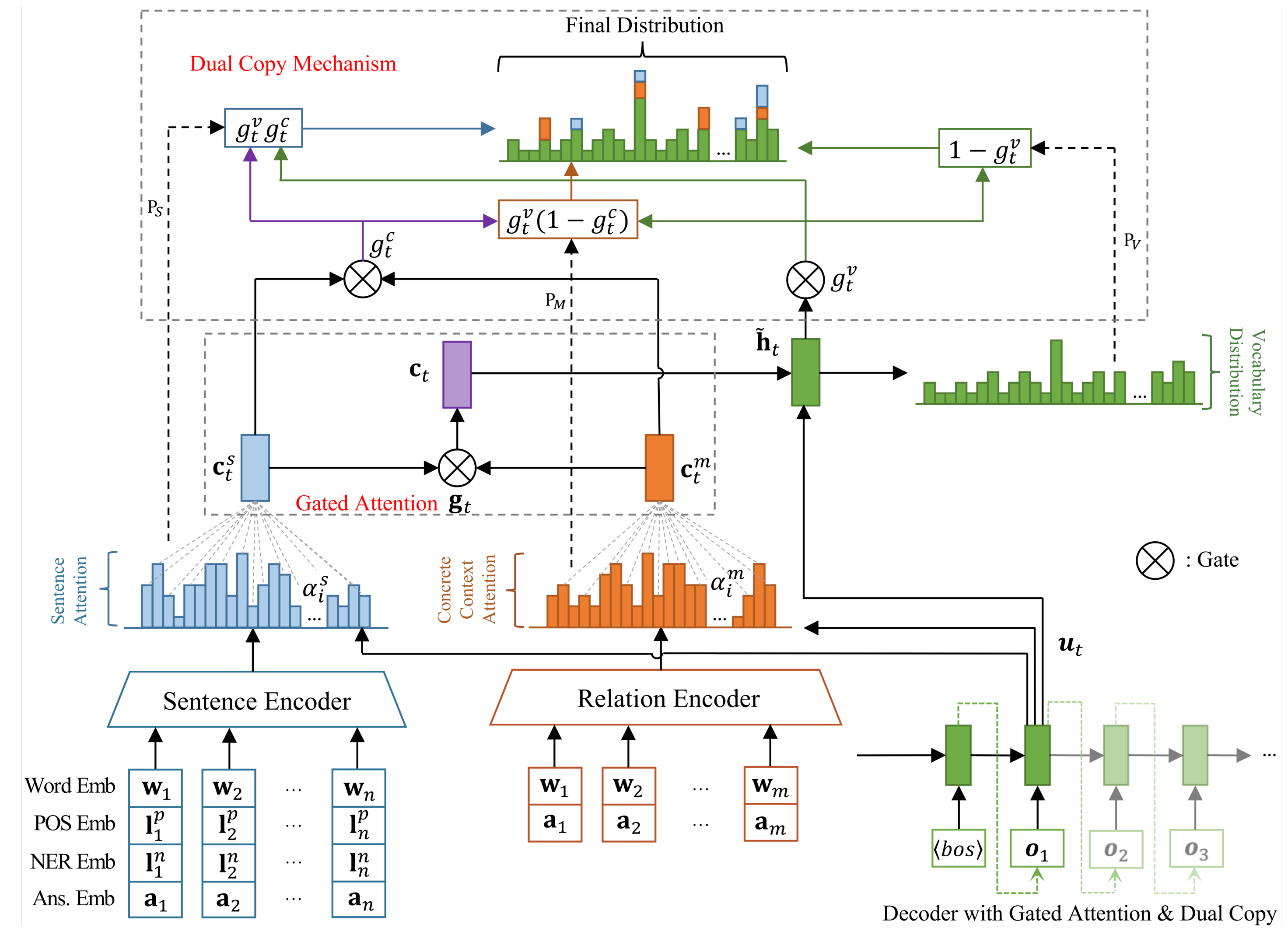
Off-the-shelf toolbox: OpenIE (Allen AI)

N -ary relation selection criterion:

- Include answer phrase;
- Get High confidence score;
- Contain maximum non-stop words.

Model

Proposed framework:



- Feature-enriched sentence encoder and relation encoder
- Gated Attention Mechanism
- Dual Copy Mechanism

Experiment Results

Main result

	Du Split (Du et al., 2017)						Zhou Split (Zhou et al., 2017)					
	B1	B2	B3	B4	MET	R-L	B1	B2	B3	B4	MET	R-L
s2s (Du et al., 2017)	43.09	25.96	17.50	12.28	16.62	39.75	-	-	-	-	-	-
NQG++ (Zhou et al., 2017)	-	-	-	-	-	-	-	-	-	13.29	-	-
M2S+cp (Song et al., 2018)	-	-	-	13.98	18.77	42.72	-	-	-	13.91	-	-
s2s+MP+GSA (Zhao et al., 2018)	43.47	28.23	20.40	15.32	19.29	43.91	44.51	29.07	21.06	15.82	19.67	44.24
Hybrid model (Sun et al., 2018)	-	-	-	-	-	-	43.02	28.14	20.51	15.64	-	-
ASs2s (Kim et al., 2019)	-	-	-	16.20	19.92	43.96	-	-	-	16.17	-	-
Our model	45.66	30.21	21.82	16.27	20.36	44.35	44.40	29.48	21.54	16.37	20.68	44.73

Performance between questions with difference answer-context proximity

(a) Evaluation results on all sentences.

	Hybrid			Our Model		
	BLEU	MET	R-L	BLEU	MET	R-L
0~10 (72.8% of #)	28.54	21.54	46.26	29.73 (4.17%)	22.03 (2.27%)	46.85 (1.28%)
>10 (27.2% of #)	20.67	16.72	37.63	22.12 (7.01%)	17.46 (4.43%)	38.47 (2.23%)

(b) Evaluation results on sentences with more than 20 words.

	Hybrid			Our Model		
	BLEU	MET	R-L	BLEU	MET	R-L
0~10 (58.3% of #)	28.00	21.03	45.37	29.11 (3.96%)	21.50 (2.21%)	45.97 (1.31%)
>10 (26.6% of #)	20.58	16.66	37.53	22.04 (7.09%)	17.38 (4.30%)	38.37 (2.24%)

Case Study

Example

Sentence: The daily mean temperature in January, the area's coldest month, is 32.6 °F (0.3 °C); however, temperatures usually drop to 10 °F (-12 °C) several times per winter and reach 50 F (10 °C) several days each winter month.

Structured Answer-relevant Relation: (The daily mean temperature in January; is; 32.6 °F (0.3 °C))

Gold Question: What is New York City's daily January mean temperature in degrees celsius ?

Baseline: What is the coldest temperature in Celsius ?

Ours: In degrees Celsius, what is the average temperature in January ?

Diverse QG

Sentence: In July 1960, NASA Deputy Administrator Hugh L. Dryden announced the Apollo program to industry representatives at a series of Space Task Group conferences.

Relation 1: (Hugh L. Dryden; [is] Deputy Administrator [of]; NASA)

Question 1: Who was the NASA Deputy Administrator in 1960 ?

Relation 2: (NASA Deputy Administrator Hugh L. Dryden; announced; the Apollo program to industry representatives at a series of Space Task Group conferences; In July 1960)

Question 2: Who announced the Apollo program to industry representatives ?