



TSGP: Two-Stage Generative Prompting for Unsupervised Commonsense Question Answering

Yueqing Sun, Yu Zhang*, Le Qi, Qi Shi

Research Center for Social Computing and Information Retrieval

Harbin Institute of Technology, Harbin, China

`{yqsun, qshi, lqi, zhangyu}@ir.hit.edu.cn`

<https://github.com/YueqingSun/TSGP>

— EMNLP 2022



Reported by Yuyang Lai



1.Introduction

2.Method

3.Experiments



Introduction

Previously —— poor generalization ability

In this paper

- knowledge generation prompts
- answer generation prompts

Question: Blue read material outside of his **comfort zone** because he wanted to gain what?

Choices: A. **new perspective** B. understanding
C. entertained D. hunger E. tired eyes

Commonsense knowledge:

I can try new things by stepping out of my comfort zone...



Answers:

growth,
new experiences
new possibilities...

Figure 1: Humans can immediately use commonsense knowledge of life to give multiple possible answers to a commonsense question. Unsupervised CSQA models should simulate human behaviour and thinking.



(a) Knowledge Generation Prompts

Instructions: Generate some knowledge about the events in the input. Examples:

Input: Kendall frightened the dogs away by yelling and waving his hands. Why did Kendall do this?

Knowledge: Parents will protect their children from dog attacks.

...

Input: Kai found one for sale online but it was too much money for her. What does Kai need to do before this?

Knowledge: people need to turn on the laptop before going online.

Input: <question>

Knowledge:

(b) Answer Generation Prompts

Instructions: For the following knowledge and question, generate the answer to the question. Examples:

Knowledge: Parents will protect their children from dog attacks.

Question: Kendall frightened the dogs away by yelling and waving his hands. Before, Kendall wanted to protect their children.

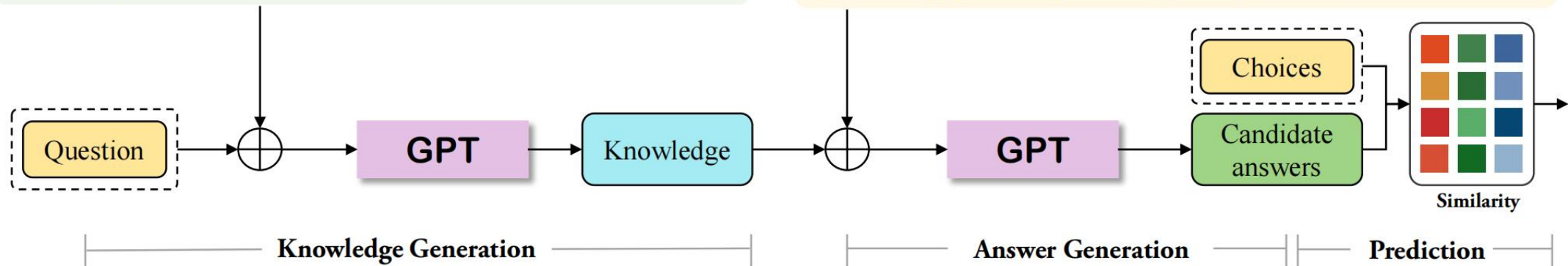
...

Knowledge: people need to turn on the laptop before going online.

Question: Kai found one for sale online but it was too much money for her. Before, Kai needed to open up her laptop.

Knowledge: <knowledge>

Question: <question> <answer prefix>





Method

(a) Knowledge Generation Prompts

Instructions: Generate some knowledge about the events in the input. Examples:

Input: Kendall frightened the dogs away by yelling and waving his hands. Why did Kendall do this?

Knowledge: Parents will protect their children from dog attacks.

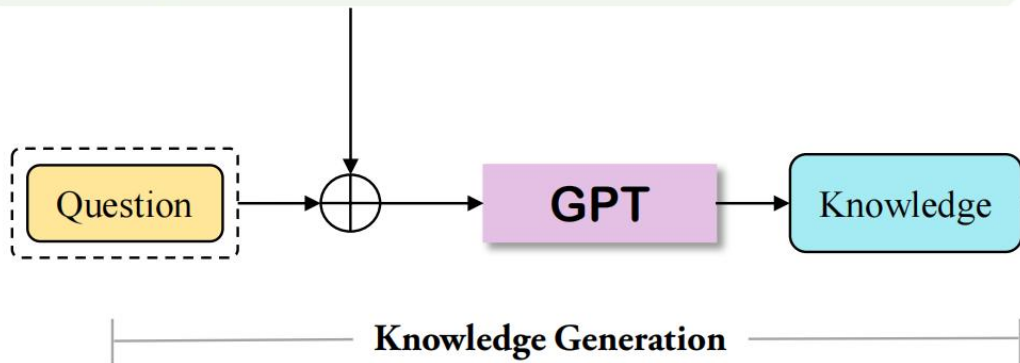
...

Input: Kai found one for sale online but it was too much money for her. What does Kai need to do before this?

Knowledge: people need to turn on the laptop before going online.

Input: <question>

Knowledge:



$$\hat{a} = \operatorname{argmax}_{a \in A} P(a|q). \quad (1)$$

$$K_q = \{k_1, k_2, \dots, k_m\}$$

$$PMI(q; k) = \log \frac{p(q|k)}{p(k)} = \log \frac{p(k|q)}{p(q)} \quad (2)$$

$$\frac{PrLM(q|k)}{PrLM(k)} = \frac{PrLM(k|q)}{PrLM(q)} \quad (3)$$

$$PMI(k; q) \stackrel{\text{def}}{=} \log \frac{PrLM(k|q)}{PrLM(q)} \quad (4)$$

$$\hat{k} = \operatorname{argmax}_{k \in K_q} PMI(k; q) \quad (5)$$

Method

(b) Answer Generation Prompts

Instructions: For the following knowledge and question, generate the answer to the question. Examples:

Knowledge: Parents will protect their children from dog attacks.

Question: Kendall frightened the dogs away by yelling and waving his hands. Before, Kendall wanted to protect their children.

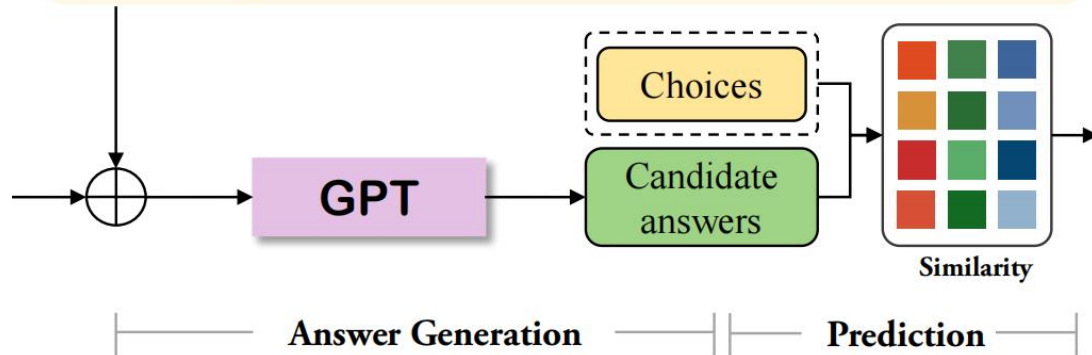
...

Knowledge: people need to turn on the laptop before going online.

Question: Kai found one for sale online but it was too much money for her. Before, Kai needed to open up her laptop.

Knowledge: <knowledge>

Question: <question> <answer prefix>



$$S_q = \{s_1, s_2, \dots, s_n\}$$

$$\hat{a} = \operatorname{argmax}_{a_i \in A} \mathcal{L}_{\text{PrLM}}(q, a_i) \quad (6)$$

$$P(a_i | S) = \frac{1}{n \cdot Z(T)} \sum_{j=1}^n \exp\left[\frac{\cos(h_{s_j}, h_{a_i})}{T}\right] \quad (7)$$



Experiments

Datasets	Models	GPT-2 Small	GPT-2 Medium	GPT-2 Large	GPT-2 XL	Published
CommonsenseQA	Baseline	29.0	29.1	32.6	32.3	–
	Self-talk	24.8	27.3	31.5	31.4	32.4 [†]
	SEQA	26.1	30.7	34.6	34.8	–
	GKP	–	–	–	–	47.3 [‡]
	TSGP (Ours)	33.3	42.2	46.8	49.1	–
OpenBookQA	Baseline	16.4	18.0	20.0	22.8	–
	Self-talk	17.4	21.0	23.8	25.4	–
	SEQA	27.6	28.6	32.0	33.4	–
	TSGP (Ours)	38.0	43.8	43.0	44.4	–
SocialIQA	Baseline	39.8	41.8	43.0	42.8	–
	Self-talk	41.2	43.3	45.3	46.2	46.2
	SEQA	44.4	44.6	46.6	47.5	47.5
	DynaGen	–	–	–	–	50.1 [◇]
	TSGP (Ours)	45.9	46.7	49.7	51.5	–



Experiments

Methods	CommonsenseQA	OpenBookQA
GPT-2 XL	32.3	22.8
+ Knowledge Generation	45.5	30.0
+ Answer Generation	43.9	42.0
+ Both (Ours)	49.1	44.4

Table 2: Ablation study on model components.

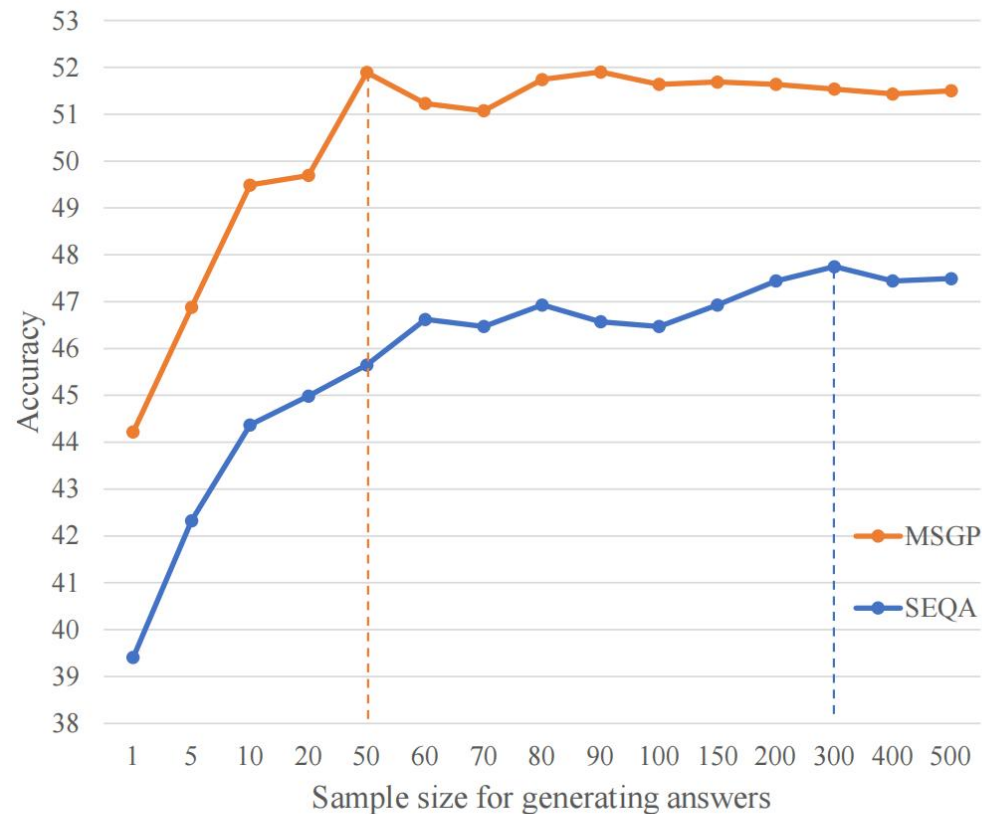


Figure 3: The effect of the number of generated answers on the accuracy. Our TSGP has the highest accuracy with an answer sample size of 50, while SEQA needs to generate more answers to achieve high accuracy.



Experiments

Model	Question / <i>Generated Knowledge</i> / Generated Answer	Predicted	Score
Baseline	Which large land mass is home to the most monkeys?	amazon basin (×)	0.35
TSGP	<i>The world's largest monkey colony lives on Madagascar.</i> “Madagascar, Africa”, “Africa”, “Madagascar, Democratic Republic of the Congo” ...	african continent (✓)	0.51
Baseline	What green area is a marmot likely to be found in?	north america (×)	0.21
TSGP	<i>Marmots live near water sources so they need plenty of vegetation around their habitat.</i> “rural area”, “country side”, “village” ...	countryside (✓)	0.34
Baseline	Where do all animals live?	zoos (×)	0.43
TSGP	<i>Animals exist on every continent except Antarctica (where they're frozen).</i> “the oceans and land”, “all over the earth”, “wherever they can find the food and shelter” ...	surface of earth (✓)	0.46

Table 3: Some examples where our model corrects the baseline (GPT2-XL) predictions. The first line of each part is the original question and the prediction result of the baseline; the second line is the knowledge statement generated by TSGP, and the third line is the pseudo-answer generated by TSGP. We show the correct answer with a check mark (✓) and the wrong answer with a wrong mark (×).



Experiments

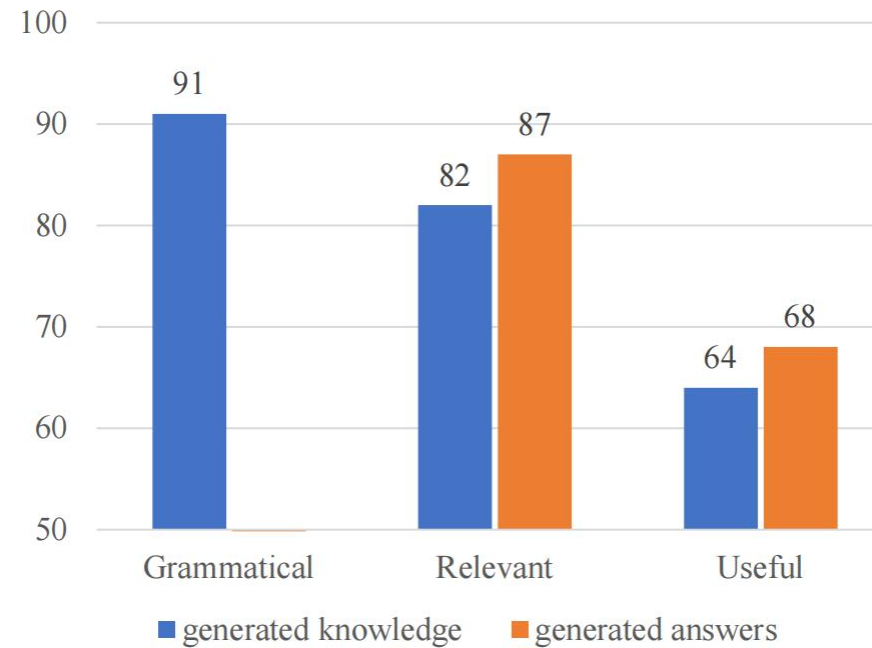


Figure 4: Human evaluation of generated knowledge and generated answers.



Thank you!