

1. (2%) There are some difference between fine-tuning and prompting. Beside fine-tuning, in-context learning enable pre-trained model to give correct prediction on many downstream tasks with a few examples but without gradient descent. Please describe:

A. How encoder-only model (Bert-series) determines the answer in a extractive question answering task?

B. How decoder-only model (GPT-series) determines the answer in a extractive question answering task?

A: By using paired extractive question answering data to fine-tuning a pre-trained model . Input only the question and paragraph ,output gets the logits of each word in paragraph and do some choosing to get answer.

B: By giving prompt and question and paragraph as input to a pre-trained model , output gets the sentence of answer.

2. (2%) The goal of this homework is to fine-tune a QA model. In this question, We wish you to try In-context learning on the same task and compare the difference between the prediction of in-context learning and fine-tuning.

A. Try to Instruct the model with different prompts, and describe your observation on at least 3 pairs of prompts comparison.

B. Paste the screenshot of your prediction file in A. to Gradescope. (There are at least 6 screenshots)

prompt: 請從最後一篇的文章中找出最後一個問題的答案  
ID, 正確答案, 模型輸出  
0, 2007年1月16日, 2007年1月16日  
1, 200公里, 200公里  
2, 廣州, 香港  
3, 2015, 1990年  
4, 文化大革命, 文化大革命  
5, 廣州, 廣州  
6, 廣州, 廣州  
7, 武昌起義, 1911年10月10日  
8, 香港, 香港和澳門  
9, 1921, 1911年10月10日  
10, 7月, 2010年7月  
11, 2008年, 2009年  
12, 從化市, 廣州的人口密度?  
13, 80, 廣州的犯罪率是廣州的犯罪率的百分  
14, 非洲黑人, 非洲黑人  
15, 普通話,

prompt: 請從文章中找出問題的答案  
ID, 正確答案, 模型輸出  
0, 2007年1月16日, 2010年  
1, 200公里, 約200公里  
2, 廣州, 廣州文章: 2010年, 廣州中央公園的語聚會受到當局的  
3, 2015, 1990  
4, 文化大革命, 文化大革命  
5, 廣州, 廣州  
6, 廣州, 廣州  
7, 武昌起義, 1911年10月10日  
8, 香港, 花都, 香, 南沙, 從化, 增城  
9, 1921, 1922年  
10, 7月, 2010年7月  
11, 2008年, 武昌起義  
12, 從化市, 越秀區  
13, 80, 廣州的社會環境如何  
14, 非洲黑人, 非洲黑人  
15, 普通話, 普通話

不提供“最後”相關的關鍵字好像對模型沒什麼關係，但是偶爾會出現奇怪的東西如右邊 2. 的答案。

prompt: 請從最後一篇的文章中找出最後一個問題的答案  
ID, 正確答案, 模型輸出  
0, 2007年1月16日, 2007年1月16日  
1, 200公里, 200公里  
2, 廣州, 香港  
3, 2015, 1990年  
4, 文化大革命, 文化大革命  
5, 廣州, 廣州  
6, 廣州, 廣州  
7, 武昌起義, 1911年10月10日  
8, 香港, 香港和澳門  
9, 1921, 1911年10月10日  
10, 7月, 2010年7月  
11, 2008年, 2009年  
12, 從化市, 廣州的人口密度?  
13, 80, 廣州的犯罪率是廣州的犯罪率是廣州的犯罪率的百分  
14, 非洲黑人, 非洲黑人  
15, 普通話,

prompt: 請閱讀前三篇文章與問題後從最後一篇的文章中找出最後一個問題的答案  
ID, 正確答案, 模型輸出  
0, 2007年1月16日, 2007年1月16日  
1, 200公里, 200公里  
2, 廣州, 廣州  
3, 2015, 1990年  
4, 文化大革命, 1975年以後, 政局改善, 政府大力支持以日用品為主的輕工業,  
5, 廣州, 廣州  
6, 廣州, 廣州  
7, 武昌起義, 廣州的總數據據廣州的總人口密度為廣州的總人口  
8, 香港, 香港和台灣  
9, 1921, 1922年  
10, 7月, 2010  
11, 2008年, 2008年  
12, 從化市, 廣州的人口密度是100倍分之一以下。  
13, 80, 80%  
14, 非洲黑人, 非洲黑人  
15, 普通話, 廣州的普通話是廣東話。

講清楚哪些文章跟答案的是甚麼任務有助模型表現。

prompt: 請從最後一篇的文章中找出最後一個問題的答案  
ID, 正確答案, 模型輸出  
0, 2007年1月16日, 2007年1月16日  
1, 200公里, 200公里  
2, 廣州, 香港  
3, 2015, 1990年  
4, 文化大革命, 文化大革命  
5, 廣州, 廣州  
6, 廣州, 廣州  
7, 武昌起義, 1911年10月10日  
8, 香港, 香港和澳門  
9, 1921, 1911年10月10日  
10, 7月, 2010年7月  
11, 2008年, 2009年  
12, 從化市, 廣州的人口密度?  
13, 80, 廣州的犯罪率是廣州的犯罪率是廣州的犯罪率的百分  
14, 非洲黑人, 非洲黑人  
15, 普通話,

prompt: 請從最後一篇找出問題的答案  
ID, 正確答案, 模型輸出  
0, 2007年1月16日, 2007年  
1, 200公里, 200公里  
2, 廣州, 廣州  
3, 2015, 1990年  
4, 文化大革命, 文化大革命  
5, 廣州, 廣州  
6, 廣州, 廣州  
7, 武昌起義, 1911年10月10日  
8, 香港, 廣州和香港  
9, 1921, 1922年  
10, 7月, 2010年7月  
11, 2008年, 2008年  
12, 從化市, 廣州的人口密度高?  
13, 80, 80%  
14, 非洲黑人, 非洲黑人  
15, 普通話, 在廣州, 許多大型公共場所, 普通話是主要語言。在某些情況

不用講“文章”模型也知道要從文章中找答案。