实训汇报0527（张勃文）

0521-0527

本周我们主要的工作是调通搜索引擎模块的模型。最终的结果是将三个ipynb文件调通并且正常运行，并设计多个测试用例测试freshprompt功能，**证明了其利用RAG解决了大模型幻觉问题。**

我们主要是为了解决大模型的幻觉问题，因此我们采用了RAG技术，也就是基于搜索引擎，根据其搜索的内容进行LLM内容的实时更新。

我们基于的是**freshllm**的论文（https://arxiv.org/abs/2310.03214），并对论文的部分代码进行了修改。为了便于测试，我们将输入文件和输出文件都保存在本地，并且设计测试用例进行测试。

遇到的困难：在调试relax和strict模块时，遇到了一些问题。如colab安装包难以安装等，最后将文件都放在本地进行，则顺利运行。

各个文件解释：

1. **freshprompt.ipynb**:主要是算法核心，一种基于RAG的大语言模型，可以根据最新的搜索到的信息进行回答。
2. **fresheval_strict**和**fresheval_relaxed.ipynb**都是评估的函数。
   「RELAXED」模式只衡量主要答案是否正确，「STRICT」模式则衡量回答中的所有说法是否都是最新的事实（即没有幻觉）。两者的区别主要在于prompt的不同。
3. **fresheval_strict_sample_evaluation_spreadsheet - freshqa.csv** 为用于「STRICT」模式的输入数据集；**fresheval_relaxed_sample_evaluation_spreadsheet - freshqa.csv** 为用于「RELAXED」模式的输入数据集；
4. **main.py**为爬虫腾讯网和搜狐网的信息代码，后期看下是否需要。@lfh

freshllm的代码示例：（见freshprompt.ipynb）

```
#@title Demonstration examples


demo_questions = [
    "What year is considered Albert Einstein's annus mirabilis?",
    "Which photographer took the most expensive photograph in the world?",
    "How many days are left until the 2023 Grammy Awards?",
    "How many years ago did the Boxing Day Tsunami happen?",
    (
        "When did Amazon become the first publicly traded company to exceed a"
        " market value of $3 trillion?"
    ),
]

concise_demo_reasonings_and_answers = [
    (
        "1905 is considered Albert Einstein's annus mirabilis, his miraculous"
        " year."
    ),
    (
        'The most expensive photograph in the world is "Le Violon d\'Ingres".'
        " The photograph was created by Man Ray."
    ),
```

```
    (
        "The 2023 Grammy Awards ceremony was held on February 5, 2023. Thus,"
        " the ceremony has already taken place."
    ),
    (
        "The disaster occurred on December 26, 2004. Thus, it happened 19 years"
        " ago."
    ),
    "Amazon's market capitalization has never exceeded $3 trillion.",
]

verbose_demo_reasonings_and_answers = [
    (
        "In the year of 1905, Albert Einstein published four groundbreaking"
        " papers that revolutionized scientific understanding of the universe."
        " Thus, scientists call 1905 Albert Einstein's annus mirabilis – his"
        " year of miracles."
    ),
    (
        "Man Ray's famed \"Le Violon d'Ingres\" became the most expensive"
        " photograph ever to sell at auction, sold for $12.4 million on May"
        " 14th, 2022 at Christie's New York. The black and white image, taken"
        " in 1924 by the American surrealist artist, transforms a woman's naked"
        " body into a violin by overlaying the picture of her back with"
        " f-holes. Thus, Man Ray is the photographer who took the most"
        " expensive photograph in the world."
    ),
    (
        "The 2023 Grammy Awards, officially known as the 65th Annual Grammy"
        " Awards ceremony, was held in Los Angeles on February 5, 2023. Thus,"
        " the event has already taken place."
    ),
    (
        "The Boxing Day Tsunami refers to the 2004 Indian Ocean earthquake and"
        " tsunami, which is one of the deadliest natural disasters in recorded"
        " history, killing an estimated 230,000 people across 14 countries. The"
        " disaster occurred on December 26, 2004, which is 19 years ago."
    ),
    (
        "Amazon's market capitalization hit a peak of roughly $1.9 trillion in"
        " July 2021. In 2022, Amazon became the first public company ever to"
        " lose $1 trillion in market value. Thus, Amazon's market value has"
        " never exceeded $3 trillion. In fact, Apple became the first publicly"
        " traded U.S. company to exceed a market value of $3 trillion in"
        " January 2022."
    ),
]

prefix = (
    f"\nanswer: As of today {current_date}, the most up-to-date and relevant"
    " information regarding this query is as follows. "
)

concise_demo_reasonings_and_answers = [
    prefix + x for x in concise_demo_reasonings_and_answers
]
verbose_demo_reasonings_and_answers = [
    prefix + x for x in verbose_demo_reasonings_and_answers
```

```
]
```

freshprompt的运行结果：（能够进行跑通）

test1：测试格莱美奖

```
#@title FreshPrompt

# @markdown ---
model_name = "gpt-3.5-turbo" #@param ["gpt-4-0125-preview", "gpt-4-turbo-preview", "gpt-4-1106-preview", "gpt-4",
check_premise = True  # @param {type:"boolean"}
# @markdown ### Ask your question here!

question = "Who is the latest artist confirmed to be performing during the 2024 Grammys telecast?"  # @param {typ
answer = call_freshprompt(model_name, question, check_premise=check_premise)

# Directly output the answer
print(answer)
```

Python

```
The latest artist confirmed to be performing during the 2024 Grammys telecast is Joni Mitchell.
```

test2：测试中国现在的chairman

```
#@title FreshPrompt

# @markdown ---
model_name = "gpt-3.5-turbo" #@param ["gpt-4-0125-preview", "gpt-4-turbo-preview", "gpt-4-1106-preview", "gpt-4", "gpt-4-0613", "gpt-4-32k", "gpt-4-
check_premise = True  # @param {type:"boolean"}
# @markdown ### Ask your question here!

# question = "Who is the latest artist confirmed to be performing during the 2024 Grammys telecast?"  # @param {type:"string"}
question = "现在的中国国家主席是胡锦涛。"
answer = call_freshprompt(model_name, question, check_premise=check_premise)

# Directly output the answer
print(answer)
```

[15] ✓ 8.6s Python

```
The information provided is outdated. Hu Jintao was the President of China from 2003 to 2013. As of my knowledge cutoff in May 2024, the current Presid
```

test3：测试美国现在的总统

```
#@title FreshPrompt

# @markdown ---
model_name = "gpt-3.5-turbo" #@param ["gpt-4-0125-preview", "gpt-4-turbo-preview",
check_premise = True  # @param {type:"boolean"}
# @markdown ### Ask your question here!

# question = "Who is the latest artist confirmed to be performing during the 2024 Gr
question = "美国总统现在是谁？"
answer = call_freshprompt(model_name, question, check_premise=check_premise)

# Directly output the answer
print(answer)
```

✓ 9.6s

```
The current President of the United States is Joe Biden.
```

结论：可以作为基于RAG的LLM系统，解决大模型幻觉问题。

下周的工作设想

将算法都写成def函数形式，整合到后端去。以及讨论一下项目还需要做的工作：比如如何凸显出对比系统这一概念；之后再结合前端形成完整系统。