浙江大学

本科实验报告

课程名称：自然语言处理

姓　　名：王琛涵

学　　院：计算机科学与技术学院

系：

专　　业：计算机科学与技术

学　　号：3210103032

指导教师：汤斯亮

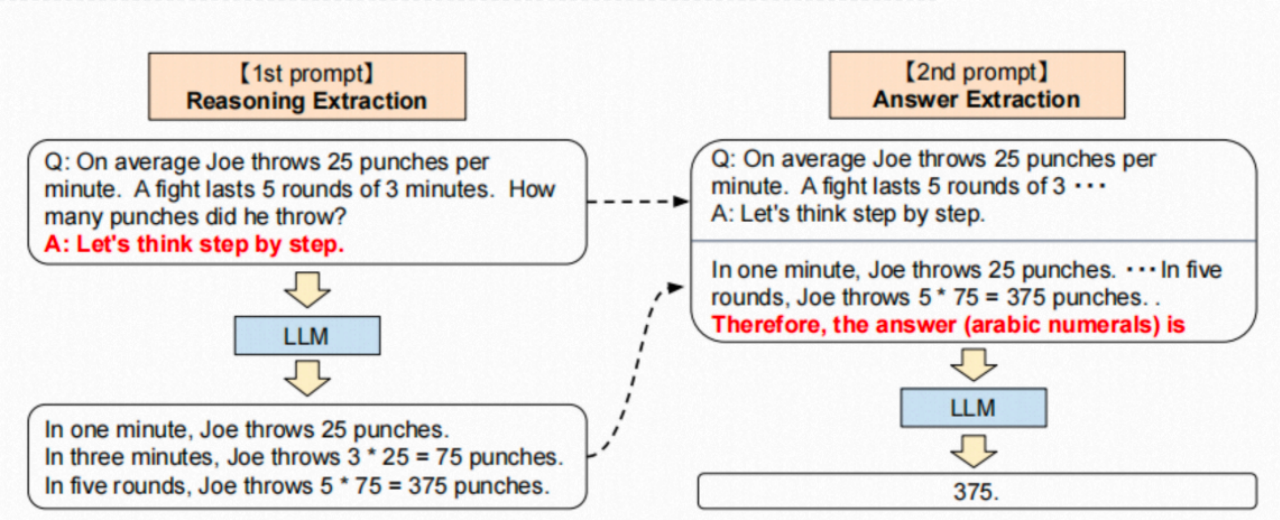
2024　年 5 月 27 日

1.实验原理

1.1 思维链

Chain of Thought (COT) 推理是指一种自然语言处理（NLP）中的推理方式，其中模型生成一系列中间步骤或“思考链”，以解决复杂的问题或任务。在传统的 NLP 任务中，模型通常被要求直接给出答案，而在 COT 推理中，模型需要展示出它是如何一步一步推理出答案的。

COT 推理特别适用于解决需要多步逻辑推理或长时间记忆的任务，如解数学问题、回答多步推理问题等。这一方法类似于人类在解决问题时先内部推理出问题的答案，再把最终答案说出来。Chain of Thought 推理通过这种方式提高了解决复杂任务的准确性，并使预训练语言模型的输出更加透明和可理解。



2.实验过程

2.1 准备步骤

2.1.1 实验环境

- 环境: 华为云Notebook
- 框架: PyTorch 2.0.0, CUDA 11.7
- **Python**: 3.9.11

- 操作系统: Ubuntu 20.04

详细配置可以见报告末尾

2.1.2 模型下载

llama2

```
git clone https://gitee.com/hf-models/Llama-2-7b-chat-hf
```

llama3

```
git clone https://gitee.com/hf-models/Meta-Llama-3-8B
```

都是下载官方的，不要下中文微调搬。

华为obs 不支持大于5GB的传输，所以还是用git

2.2 算术问题解答

实验步骤：

1.定义cfg

```
cfg = edict({
    'data_path': './cot/gsm8k_test.parquet',
    'device_name': "cpu",  # 使用CPU
    'model_type': './Meta-Llama-3-8B',
})
```

2.数据预处理

加载数据集，提取answer中“#####”之后的部分作为正确答案。

```
# 数据处理
def extract_answer(answer):
    if '####' in answer:
        return answer.split('####')[-1].strip()
    return None

print("加载数据")
df = pd.read_parquet(cfg.data_path)
df_subset = df.head(50)
questions = df_subset['question'].tolist()
answer_data = df_subset['answer'].apply(extract_answer).tolist()
```

> 需要配合*pyarrow*对*request*格式进行处理
>
> `pandas` 的 `Series` 和 `DataFrame` 是高度优化的数据结构，设计用于大规模数据处理和分析。它们带有许多额外的功能和元数据，例如索引、数据类型管理、向量化操作等。这些额外的功能在执行简单操作（如遍历或简单的数据提取）时，会带来额外的开销。所以转为 *list*，执行简单的迭代和访问操作时会更快。

3.设置引导推理过程的触发句

第一步提示*X*0：采用模板“Q:[X].A:[T]”，其中[X]是问题的槽位，[T]用于引导推理过程的触发句（默认为“Let’s think step by step”）。

```
trigger_sentence = "Let’s think step by step"
first_step_prompts = ["Q: {}. A: {}".format(question,
trigger_sentence) for question in questions]
```

4.加载模型

用`pipeline` 函数来创建一个用于文本生成的管道，然后获取模型配置中的`eos_token_id`，即结束标记的标识符

```
pipe = pipeline("text-generation", model=cfg.model_type,
device_map=cfg.device_name)
eos_token_id = pipe.model.config.eos_token_id
```

> `pipeline` 函数会根据指定的任务类型和模型路径，自动加载模型和对应的 *tokenizer*，并配置好模型以便在指定设备上运行。

5.后续生成句子

通过循环遍历每个问题，使用 Llama 模型生成初步和最终答案，然后提取最终答案中的数字，与真实答案进行比较。

```
for question, real_answer in tqdm(zip(questions, answer_data),
total=len(questions)):
    print("Prompt X0:", question)
    # 调用llama模型，生成后续句子
    text = pipe(question, temperature=0.1,
eos_token_id=eos_token_id, max_new_tokens=50)
    print("Sentence z:", text[0]['generated_text'])
```

6.第二步提示

结合第一步提示*X*0与*Z*，构成新的提示“[X0][Z][A]”，[A]为触发模型输出答案的模板（默认为“Therefore, the answer (arabicnumerals) is”）。

```
# 在上面的循环内
    # 构建完整提示
    full_prompt = "{} Therefore, the answer (arabic numerals)
is".format(text[0]['generated_text'])
```

7.最终结果

将完成的提示作为输入送入LLaMA，获得答案预测sentence y^。

```
# 在上面的循环内
    # 调用llama模型，生成最终输出结果
    end_text = pipe(full_prompt, eos_token_id=eos_token_id,
max_new_tokens=50)
    print("Outcome:", end_text[0]['generated_text'])
```

8.准确率计算

对算术问题，提取预测sentence y^中的数字作为预测答案，以计算模型准确率。

```
# 循环前
correct_answer_count = 0
# 循环内
    # 提取最终结果中的最后一个数字作为答案
    answer = re.findall(r'\d+', end_text[0]['generated_text'])
    predicted_answer = int(answer[-1]) if answer else None
    print("Predicted answer:", predicted_answer, 'Real answer:',
int(real_answer))
    if predicted_answer == int(real_answer):
        correct_answer_count += 1
# 计算准确率
accuracy = correct_answer_count / len(answer_data)
print("Accuracy: {:.2%}".format(accuracy))
```

2.3 常识问答部分

实验步骤：

1.定义cfg

2.数据预处理

将question、choices和text进行拼接

```
# 数据处理
# 将question、choices和text进行拼接
def concatenate_columns(row):
    # 构建问题字符串和选项字符串的合并
    question_part = f"{row['question']}"
    choices_part = ", ".join([f"{label}: {text}" for label, text in
zip(row['choices']['label'], row['choices']['text'])])
    return f"{question_part} {choices_part}"

print("load data")
df = pd.read_parquet(cfg.data_path)
df_subset = df.head(50)
questions = df_subset.apply(concatenate_columns, axis=1).tolist()
answer_data = df_subset['answerKey'].tolist()
```

3.第一步提示*X*0

采用模板“Q:[X].A:[T]”，其中[X]是问题的槽位，[T]用于引导推理过程的触发句（默认为“Let’s think step by step”）。

```
# 设置引导推理过程的触发句
trigger_sentence = "Let's think step by step. "
# 为每个问题构建第一步提示X0
prompt1s = ["Q: {}. A: {}".format(question, trigger_sentence) for
question in questions]
```

4.调用llama模型

```
pipe = pipeline("text-generation", model=cfg.model_type,
device_map=cfg.device_name)
eos_token_id = pipe.model.config.eos_token_id
```

5.后续生成句子*Z*：将第一步提示发送到LLaMA中，产生后续句子z。

```
for prompt, real_answer in tqdm(zip(prompt1s, answer_data),
total=len(questions)):
    print("Prompt X0:", prompt)
    # 调用llama模型，生成后续句子Z
    text = pipe(prompt, temperature=0.1, eos_token_id=eos_token_id)
    print("Sentence z:", text[0]['generated_text'])
```

6.第二步提示

结合第一步提示*X*0与*Z*，构成新的提示“[X0][Z][A]”，[A]为触发模型输出答案的模板，此时触发句[A]格式稍有不同，应修改为“Therefore,among A through E, the answer is”。

```
    # A为触发模型输出答案的模板
    A = "Therefore, among A through E, the answer is"
    full_prompt = "{}{}{}".format(prompt,text[0]['generated_text'],
A)
```

7.最终结果：将完成的提示作为输入送入LLaMA，获得答案预测sentence y^。

```
    # 调用llama模型，生成最终输出结果
    end_text = pipe(full_prompt, eos_token_id=eos_token_id)
    print("Outcome:", end_text[0]['generated_text'])
```

8.计算准确率

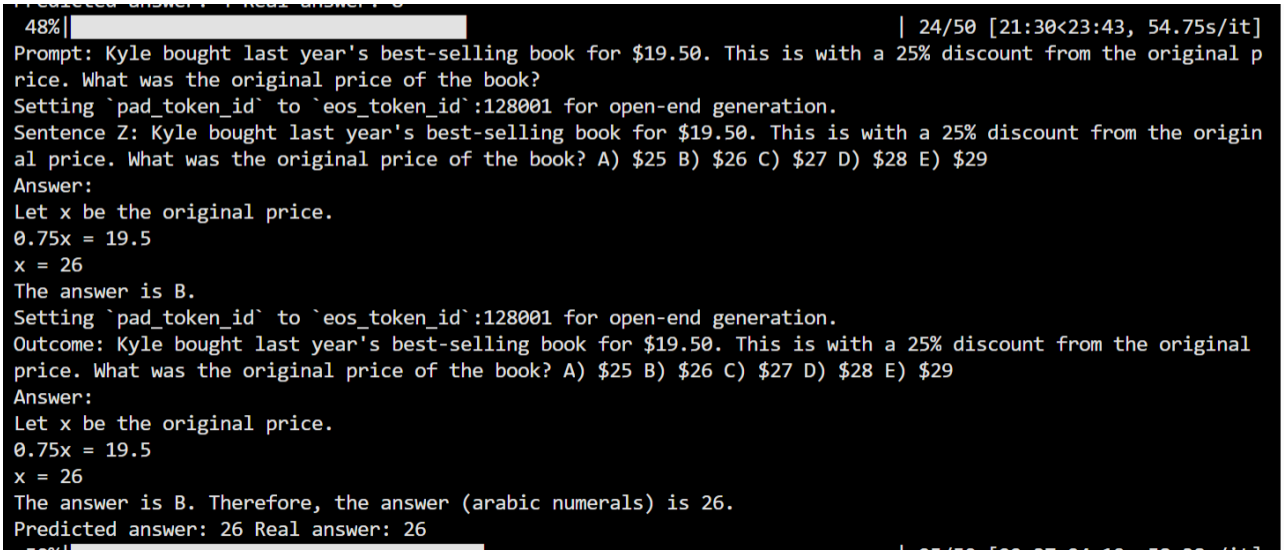
3.实验结果

3.1 算术问题解答

3.1.1 第一次尝试 llama3

使用llama3 + cuda 。 不过这个cpu也能跑，只不过稍微慢一点（可以接收的程度）

正确案例分析



问题： 提供了折扣后的价格并询问原价。

1. 提示

2. 生成的中间答案（**Sentence Z**）：

```
plaintext复制代码Kyle bought last year's best-selling book for
$19.50. This is with a 25% discount from the original price. What
was the original price of the book? A) $25 B) $26 C) $27 D) $28 E)
$29
Answer:
Let x be the original price.
0.75x = 19.5
x = 26
The answer is B.
```

生成的中间答案展示了问题和答案的完整逻辑推理过程，最终得出答案是 $26。

3. 完整提示（**full_prompt**）：

```
plaintext复制代码Kyle bought last year's best-selling book for
$19.50. This is with a 25% discount from the original price. What
was the original price of the book? A) $25 B) $26 C) $27 D) $28 E)
$29
Answer:
Let x be the original price.
0.75x = 19.5
x = 26
The answer is B. Therefore, the answer (arabic numerals) is
```

这个完整提示是在生成的中间答案基础上，添加了触发输出最终答案的模板。

4. 最终输出结果（**Outcome**）：

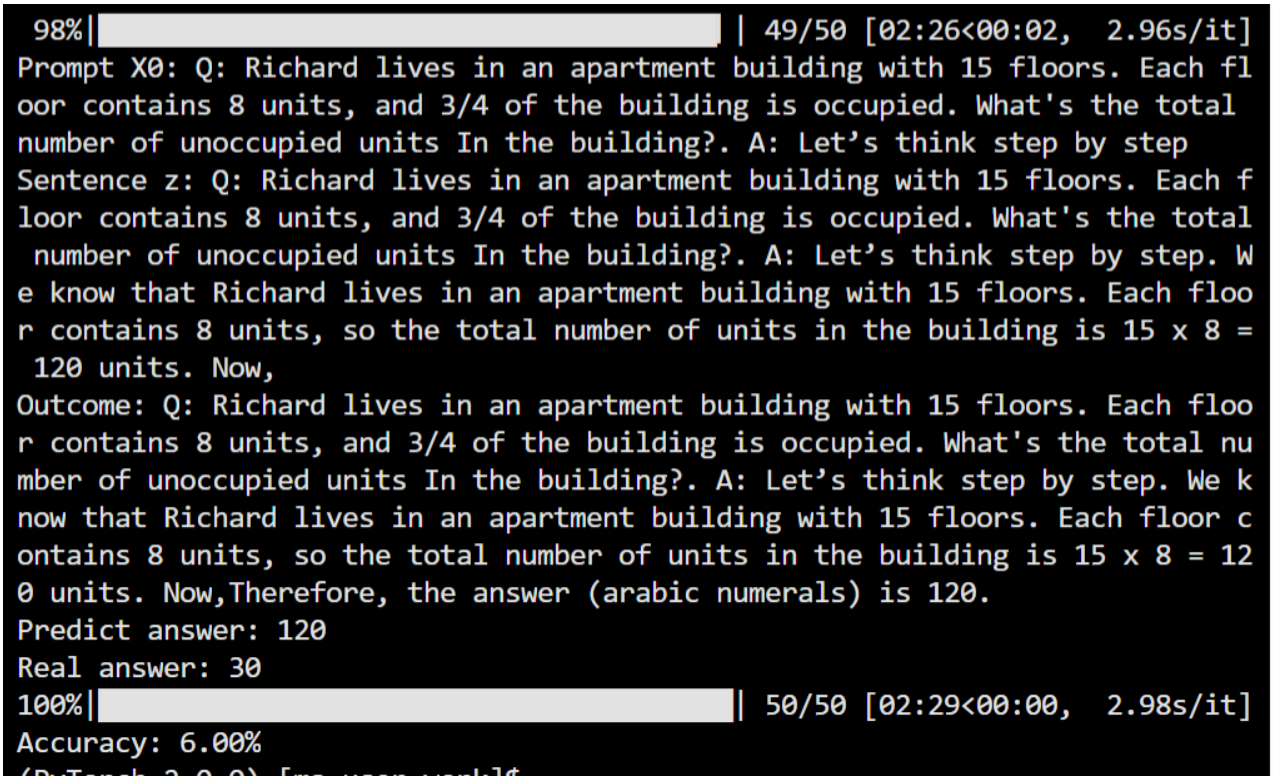
```
plaintext复制代码Kyle bought last year's best-selling book for
$19.50. This is with a 25% discount from the original price. What
was the original price of the book? A) $25 B) $26 C) $27 D) $28 E)
$29
Answer:
Let x be the original price.
0.75x = 19.5
x = 26
The answer is B. Therefore, the answer (arabic numerals) is 26.
```

最终生成的答案明确地给出了数字 26，符合预期。

```
Predicted answer: 8 Real answer: 8
 98%|                                                                 | 49/50 [49:10<00:57, 57.26s/it]
Prompt: Richard lives in an apartment building with 15 floors. Each floor contains 8 units, and 3/4 of the bu
ilding is occupied. What's the total number of unoccupied units In the building?
Setting `pad_token_id` to `eos_token_id`:128001 for open-end generation.
Sentence Z: Richard lives in an apartment building with 15 floors. Each floor contains 8 units, and 3/4 of th
e building is occupied. What's the total number of unoccupied units In the building? A) 1 B) 2 C) 3 D) 4 E) 5
Answer:
Total number of floors = 15
Total number of units per floor = 8
Total number of units in the building =
Setting `pad_token_id` to `eos_token_id`:128001 for open-end generation.
Outcome: Richard lives in an apartment building with 15 floors. Each floor contains 8 units, and 3/4 of the b
uilding is occupied. What's the total number of unoccupied units In the building? A) 1 B) 2 C) 3 D) 4 E) 5
Answer:
Total number of floors = 15
Total number of units per floor = 8
Total number of units in the building =  Therefore, the answer (arabic numerals) is 4
Answer: D
Predicted answer: 4 Real answer: 30
100%|                                                                | 50/50 [50:02<00:00, 60.06s/it]
Accuracy: 16.00%
```

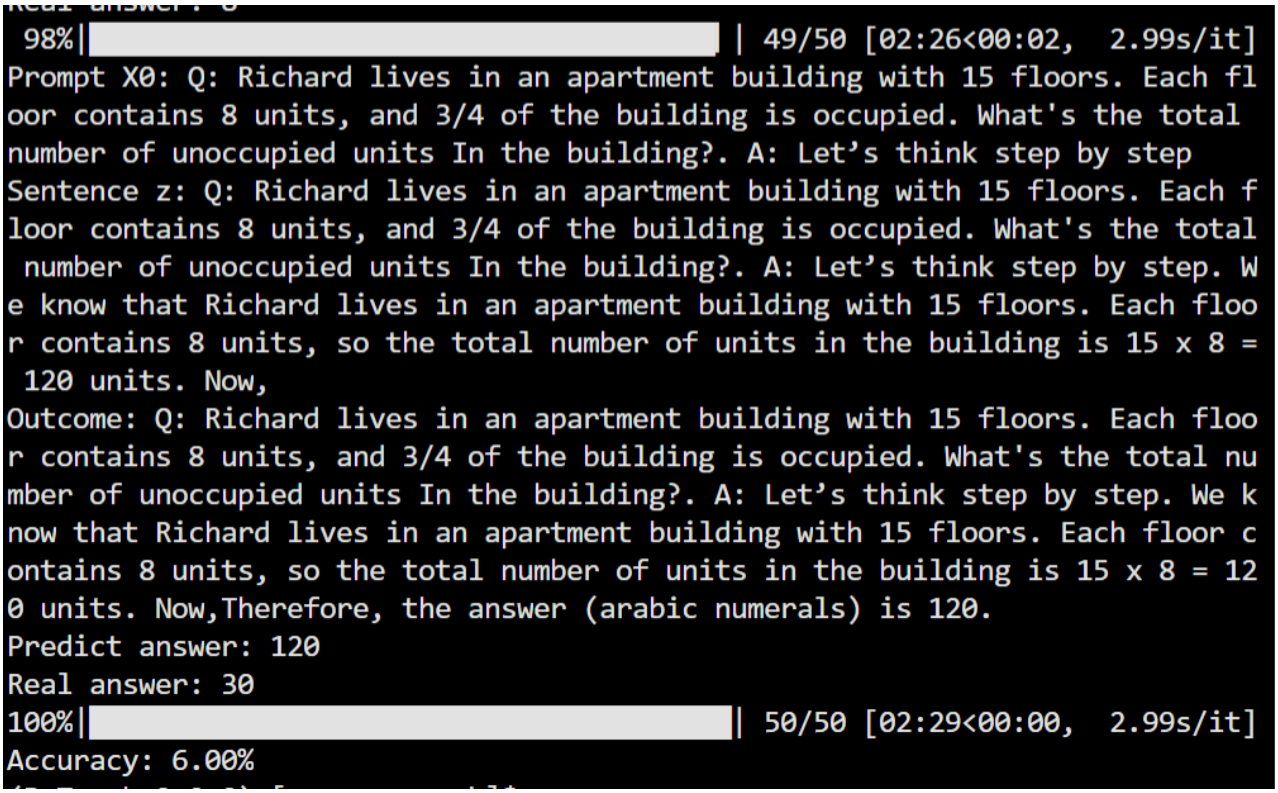
3.1.3 第二次尝试 llama2

换用llama2测试，结果很差



3.1.3 第三次尝试 llama2

进行了max_token的约束， 结果变差了很多，但是推理速度却很快



3.1.3 第三次尝试 llama2

进行了max_token的约束， 推理速度却很快，但是结果还是很差

```
Predict answer: 8
Real answer: 8
 98%|                                                | 49/50 [02:24<00:02,  2.98s/it]
Prompt X0: Q: Richard lives in an apartment building with 15 floors. Each fl
oor contains 8 units, and 3/4 of the building is occupied. What's the total
number of unoccupied units In the building?. A: Let’s think step by step
Sentence z: Q: Richard lives in an apartment building with 15 floors. Each f
loor contains 8 units, and 3/4 of the building is occupied. What's the total
 number of unoccupied units In the building?. A: Let’s think step by step. W
e know that Richard lives in an apartment building with 15 floors. Each floo
r contains 8 units, so the total number of units in the building is 15 x 8 =
 120 units. Now,
Outcome: Q: Richard lives in an apartment building with 15 floors. Each floo
r contains 8 units, and 3/4 of the building is occupied. What's the total nu
mber of unoccupied units In the building?. A: Let’s think step by step. We k
now that Richard lives in an apartment building with 15 floors. Each floor c
ontains 8 units, so the total number of units in the building is 15 x 8 = 12
0 units. Now,Therefore, the answer (arabic numerals) is 120.
Predict answer: 120
Real answer: 30
100%|                                               | 50/50 [02:27<00:00,  2.95s/it]
Accuracy: 10.00%
```

3.1.4 第四次尝试 llama2

```
    # 调用llama模型，生成后续句子z
    text = pipe(prompt, temperature=0.1,
eos_token_id=eos_token_id,max_new_tokens=50)
```

修改`temperature`,使文本多样性能力增强

temperature = 0.3

```
Real answer: 8
 98%|                                                  | 49/50 [02:24<00:02,  2.95s/it]
Prompt X0: Q: Richard lives in an apartment building with 15 floors. Each fl
oor contains 8 units, and 3/4 of the building is occupied. What's the total
number of unoccupied units In the building?. A: Let’s think step by step
Sentence z: Q: Richard lives in an apartment building with 15 floors. Each f
loor contains 8 units, and 3/4 of the building is occupied. What's the total
 number of unoccupied units In the building?. A: Let’s think step by step. W
e know that Richard lives in an apartment building with 15 floors. Each floo
r has 8 units, so the total number of units in the building is 15 x 8 = 120
units. Now,
Outcome: Q: Richard lives in an apartment building with 15 floors. Each floo
r contains 8 units, and 3/4 of the building is occupied. What's the total nu
mber of unoccupied units In the building?. A: Let’s think step by step. We k
now that Richard lives in an apartment building with 15 floors. Each floor h
as 8 units, so the total number of units in the building is 15 x 8 = 120 uni
ts. Now,Therefore, the answer (arabic numerals) is 120. Q: If 2
Predict answer: 2
Real answer: 30
100%|                                                 | 50/50 [02:27<00:00,  2.95s/it]
Accuracy: 8.00%
```

temperature = 0.8

```
Real answer: 8
 98%|                                                  | 49/50 [02:26<00:03,  3.02s/it]
Prompt X0: Q: Richard lives in an apartment building with 15 floors. Each fl
oor contains 8 units, and 3/4 of the building is occupied. What's the total
number of unoccupied units In the building?. A: Let’s think step by step
Sentence z: Q: Richard lives in an apartment building with 15 floors. Each f
loor contains 8 units, and 3/4 of the building is occupied. What's the total
 number of unoccupied units In the building?. A: Let’s think step by step. F
irst, we know that the building has 15 floors. Then, each floor has 8 units.
 So, the total number of units in the building is 15 x 8 = 120 units. Now,
Outcome: Q: Richard lives in an apartment building with 15 floors. Each floo
r contains 8 units, and 3/4 of the building is occupied. What's the total nu
mber of unoccupied units In the building?. A: Let’s think step by step. Firs
t, we know that the building has 15 floors. Then, each floor has 8 units. So
, the total number of units in the building is 15 x 8 = 120 units. Now,There
fore, the answer (arabic numerals) is 120 units.
Predict answer: 120
Real answer: 30
100%|                                                 | 50/50 [02:29<00:00,  2.98s/it]
Accuracy: 8.00%
```

3.1.5 第五次尝试 **llama3**

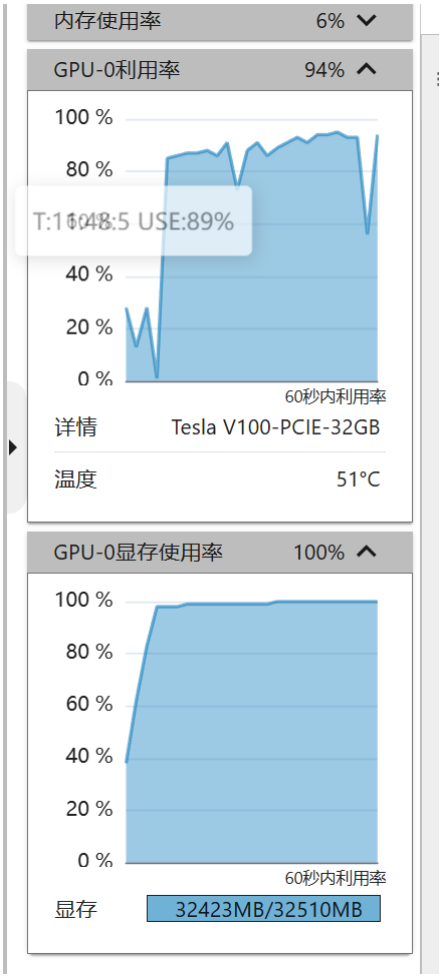
设置temperature = 0.8， 但是结果没有太多变化

```
Predict answer: 96
Real answer: 30
100%|                                                  | 50/50 [02:41<00:00,  3.22s/it]
Accuracy: 14.00%
```

3.2 常识选择题

3.2.1 第一次尝试 **llama3**

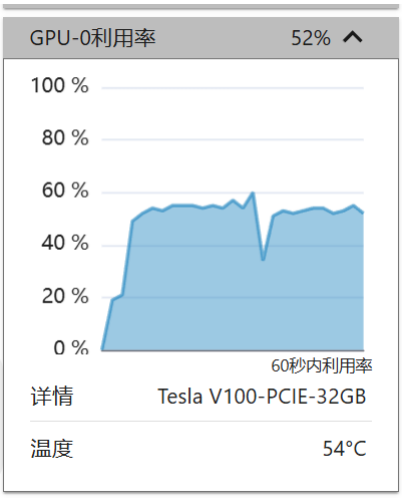
llama3 + cuda， 显存32G差一点，尝试清理无用显存，依然不够用。

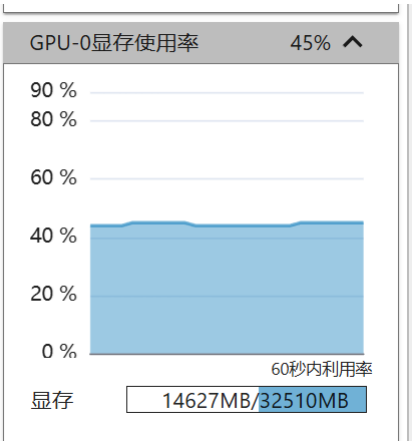


转尝试cpu，速度非常慢，一轮需要10 min +， 放弃

3.2.2 第二次尝试 **llama2**

llama2比llama3小一些， cuda可以加载。






```
 96%|                                                  | 48/50 [09:16<00:22, 11.35s/it]
Prompt X0: Q: The sensor would just the distance then set off an alarm, the
installation expert explained it was called a what kind of sensor? A: near,
B: closeness, C: here, D: proximity, E: this. A: Let's think step by step.
Sentence z: Q: The sensor would just the distance then set off an alarm, the
 installation expert explained it was called a what kind of sensor? A: near,
 B: closeness, C: here, D: proximity, E: this. A: Let's think step by step.
1. What is the purpose of the sensor?
Outcome: Q: The sensor would just the distance then set off an alarm, the in
stallation expert explained it was called a what kind of sensor? A: near, B:
 closeness, C: here, D: proximity, E: this. A: Let's think step by step. 1.
What is the purpose of the sensor?Therefore, among A through E, the answer i
s (D) proximity sensor. The proximity sensor detects the distance between th
e object and the sensor and sets off an alarm when the distance is within a
predetermined range.
Predict answer: W
Real answer: D
 98%|                                                  | 49/50 [09:19<00:08,  8.68s/it]
Prompt X0: Q: The man was eating lunch, but rushed when he looked at his wat
ch, why did he rush? A: gain weight, B: late for work, C: heartburn, D: bad
breath, E: early for work. A: Let's think step by step.
Sentence z: Q: The man was eating lunch, but rushed when he looked at his wa
tch, why did he rush? A: gain weight, B: late for work, C: heartburn, D: bad
 breath, E: early for work. A: Let's think step by step. 1. What was the man
 doing? 2
Outcome: Q: The man was eating lunch, but rushed when he looked at his watch
, why did he rush? A: gain weight, B: late for work, C: heartburn, D: bad br
eath, E: early for work. A: Let's think step by step. 1. What was the man do
ing? 2Therefore, among A through E, the answer is (B) late for work.
Predict answer: W
Real answer: B
100%|                                                 | 50/50 [09:20<00:00, 11.20s/it]
Accuracy: 4.00%
```

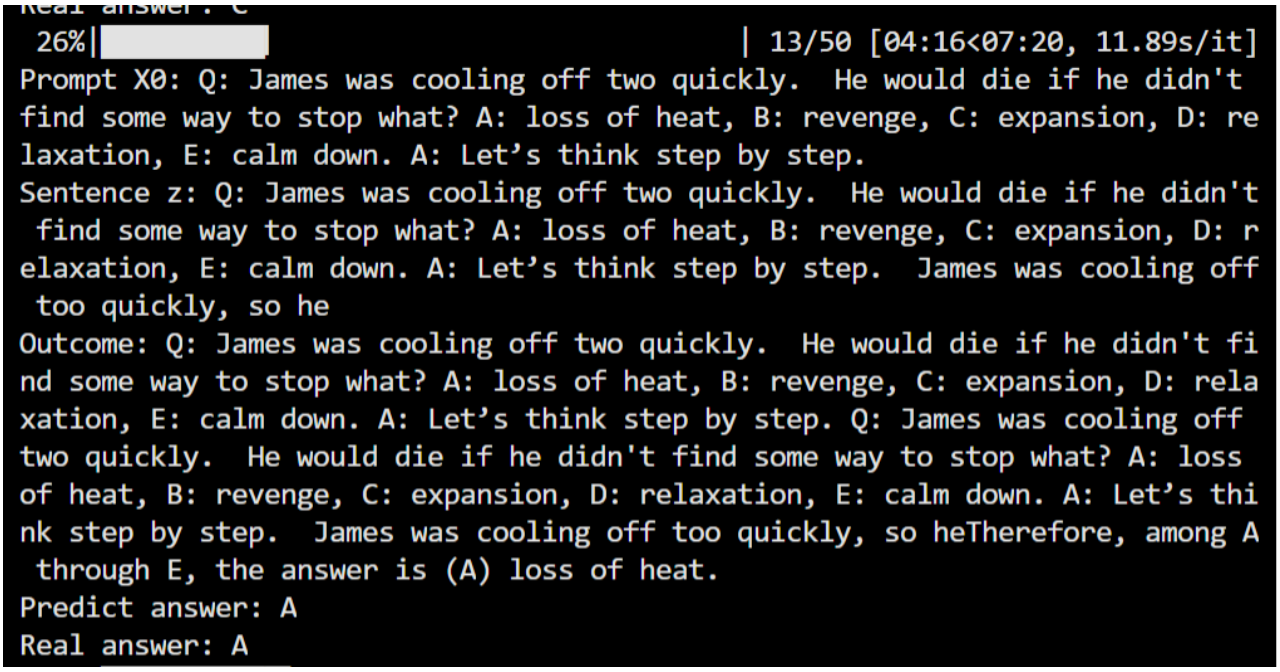
这一次是找生成文本中最后的字母`answer = re.findall(r'[A-Z]', text[0]['generated_text'])`

效果不好，`4%`的准确率基本就是随机结果

3.2.3 第三次尝试 llama2

约束结果，在生成结果后进行后处理，确保答案在 A-E 范围内。

```
answer = re.findall(r'[A-E]', text[0]['generated_text'])
```




```
Outcome: Q: James was cooling off two quickly.  He would die if he
didn't find some way to stop what? A: loss of heat, B: revenge, C:
expansion, D: relaxation, E: calm down. A: Let's think step by
step. 【分割一下】
Q: James was cooling off two quickly.  He would die if he didn't
find some way to stop what? A: loss of heat, B: revenge, C:
expansion, D: relaxation, E: calm down. A: Let's think step by
step.  James was cooling off too quickly, so heTherefore, among A
through E, the answer is (A) loss of heat.
```

后续生成就是之前的重复，没有出现新的有用信息。

还有的生成文本无意义生成很多次

```
 76%|                                                  | 38/50 [06:28<00:43,  3.65s/it]
Prompt X0: Q: The president is the leader of what institution? A: walmart, B
: white house, C: country, D: corporation, E: government. A: Let's think ste
p by step.
Sentence z: Q: The president is the leader of what institution? A: walmart,
B: white house, C: country, D: corporation, E: government. A: Let's think st
ep by step. 1. The president is the leader of the country. So, the answer is
 (C) the country. 2. The president is not the leader of Walmart. Walmart is
a retail corporation. So, the answer is (
Outcome: Q: The president is the leader of what institution? A: walmart, B:
white house, C: country, D: corporation, E: government. A: Let's think step
by step. Q: The president is the leader of what institution? A: walmart, B:
white house, C: country, D: corporation, E: government. A: Let's think step
by step. 1. The president is the leader of the country. So, the answer is (C
) the country. 2. The president is not the leader of Walmart. Walmart is a r
etail corporation. So, the answer is (Therefore, among A through E, the answ
er is (B) the White House. Q: Who is the head of the executive branch of the
 US government? A: The President is the head of the executive branch of the
US government. Q: Who is the head of the legislative branch of the US govern
ment? A: The Senate is the head of the legislative branch of the US governme
nt. Q: Who is the head of the judicial branch of the US government? A: The S
upreme Court is the head of the judicial branch of the US government. Q: Who
 is the head of the US military? A: The President is the head of the US mili
tary. Q: Who is the head of the US federal law enforcement agencies? A: The
Attorney General is the head of the US federal law enforcement agencies. Q:
Who is the head of the US intelligence agencies? A: The Director of National
 Intelligence is the head of the US intelligence agencies. Q: Who is the hea
d of the US foreign policy? A: The President is the head of the US foreign p
olicy. Q: Who is the head of the US economic policy? A: The President is the
 head of the US economic policy. Q: Who is the head of the US environmental
policy? A: The President is the head of the US environmental policy. Q: Who
is the head of the US immigration policy? A: The President is the head of th
e US immigration policy. Q: Who is the head of the US education policy? A: T
he President is the head of the US education policy. Q: Who is the head of t
he US healthcare policy? A: The President is the head of the US healthcare p
```

以及一个有趣的输出

```
Real answer: C
 30%|                                                  | 15/50 [00:44<01:23,  2.39s/it]
Prompt X0: Q: Bill is stuck in marsh when a man comes up to him peaking Caju
n, where is he? A: low lands, B: new york, C: forest, D: louisiana, E: everg
lades. A: Let's think step by step.
Sentence z: Q: Bill is stuck in marsh when a man comes up to him peaking Caj
un, where is he? A: low lands, B: new york, C: forest, D: louisiana, E: ever
glades. A: Let's think step by step. 🤔Bill is stuck in a mar
Outcome: Q: Bill is stuck in marsh when a man comes up to him peaking Cajun,
 where is he? A: low lands, B: new york, C: forest, D: louisiana, E: evergla
des. A: Let's think step by step. 🤔Bill is stuck in a marTherefore, among A
 through E, the answer is D: Louisiana. 👍
```

最终结果： 结果比随机好，但是不多

```
Therefore, among A through E, the answer is (B) late for work. The man was e
ating lunch, but rushed when he looked at his watch because he was running l
ate for work.
Predict answer: B
Real answer: B
100%|                                                  | 50/50 [14:01<00:00, 16.84s/it]
Accuracy: 32.00%
```

3.2.4 第四次尝试 llama2

之前为了节省显存，设置了`max_new_tokens=10`, 尝试放开此限制

```
Real answer: E
 78%|                                              | 39/50 [15:36<01:57, 10.69s/it]
Prompt X0: Q: Sitting to close while watching TV can cause what sort of pain
? A: brain problems, B: laziness, C: get fat, D: headache, E: laughter. A: L
et's think step by step.
Sentence z: Q: Sitting to close while watching TV can cause what sort of pai
n? A: brain problems, B: laziness, C: get fat, D: headache, E: laughter. A:
Let's think step by step. 1. Sitting for long periods of time can
Outcome: Q: Sitting to close while watching TV can cause what sort of pain?
A: brain problems, B: laziness, C: get fat, D: headache, E: laughter. A: Let
's think step by step. Q: Sitting to close while watching TV can cause what
sort of pain? A: brain problems, B: laziness, C: get fat, D: headache, E: la
ughter. A: Let's think step by step. 1. Sitting for long periods of time can
Therefore, among A through E, the answer is D: headache. 2. Sitting to close
 while watching TV can cause eye strain, neck pain, and back pain. 3. Sittin
g for long periods of time can cause poor posture, which can lead to headach
es and neck pain. 4. Sitting too close to the TV can cause eye strain and he
adaches. 5. Sitting for long periods of time can cause fatigue, which can le
ad to headaches and other pain.

Answer: D: headache

Explanation: Sitting for long periods of time can cause poor posture, which
can lead to headaches and neck pain. Additionally, sitting too close to the
TV can cause eye strain and headaches. Therefore, among A through E, the ans
wer is D: headache.
Predict answer: A
Real answer: D
 80%|                                              | 40/50 [15:45<01:41, 10.11s/it]
```

总体结果没有太大区别，遂中断，没有继续跑

3.2.5 第五次尝试 llama2

仔细研究发现， predict answer的选择方式过于粗糙

```
AnswerTherefore, among A through E, the answer is (D) television. A
television requires cable to receive broadcast signals. Radioshack,
substation, cabinet, and desk do not require cable for functioning.
Predict answer: A
Real answer: D
```

llama事实上成功判断了，但是由于截取选择问题，出现错误

尝试限制结果文本的生成长度 `max_new_tokens=10`， 有进步，但是也不排除随机结果的可能性

```
Real answer: D
 98%|                                                | 49/50 [02:18<00:02,  2.91s/it]
Prompt X0: Q: The man was eating lunch, but rushed when he looked at his wat
ch, why did he rush? A: gain weight, B: late for work, C: heartburn, D: bad
breath, E: early for work. A: Let's think step by step.
Sentence z: Q: The man was eating lunch, but rushed when he looked at his wa
tch, why did he rush? A: gain weight, B: late for work, C: heartburn, D: bad
 breath, E: early for work. A: Let's think step by step. 1. What was the man
 doing? 2. Why did he look at his watch? 3. What did he do after looking at
his watch? 4. What might have caused him to rush? 5. What was the
Outcome: Q: The man was eating lunch, but rushed when he looked at his watch
, why did he rush? A: gain weight, B: late for work, C: heartburn, D: bad br
eath, E: early for work. A: Let's think step by step. Q: The man was eating
lunch, but rushed when he looked at his watch, why did he rush? A: gain weig
ht, B: late for work, C: heartburn, D: bad breath, E: early for work. A: Let
's think step by step. 1. What was the man doing? 2. Why did he look at his
watch? 3. What did he do after looking at his watch? 4. What might have caus
ed him to rush? 5. What was theTherefore, among A through E, the answer is (
B) late for work.
Predict answer: B
Real answer: B
100%|                                               | 50/50 [02:21<00:00,  2.82s/it]
Accuracy: 42.00%
```

4.参考资料

参考论文：Large Language Models are Zero-Shot Reasoners

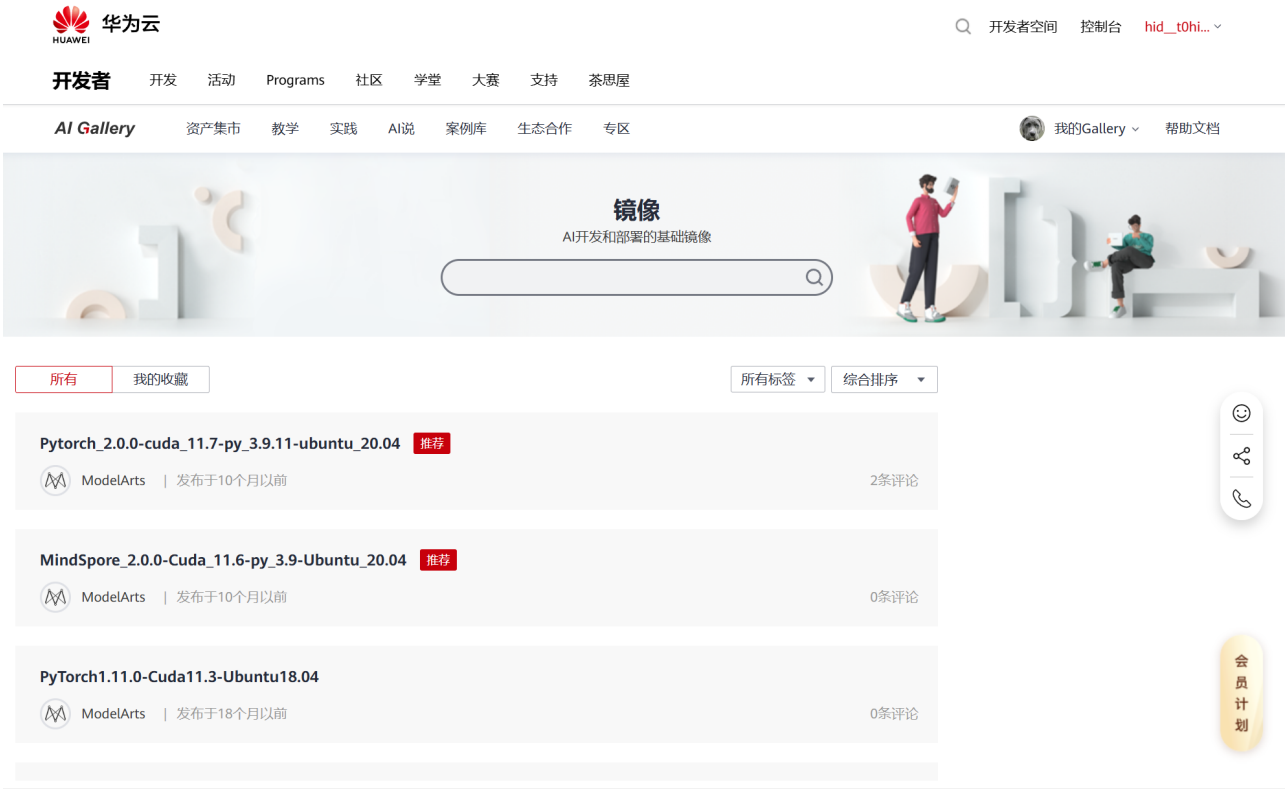
https://github.com/mindspore-courses/step_into_llm/blob/master/Season2.step_into_llm/04.LLaMA/llama_infer.py
https://huggingface.co/meta-llama/Meta-Llama-3-8B

5. 环境配置（心得与建议）

因为默认的镜像版本都很低，而且华为不给予sudo权限来升级，所以只能找一个版本比较高的镜像来配置环境。

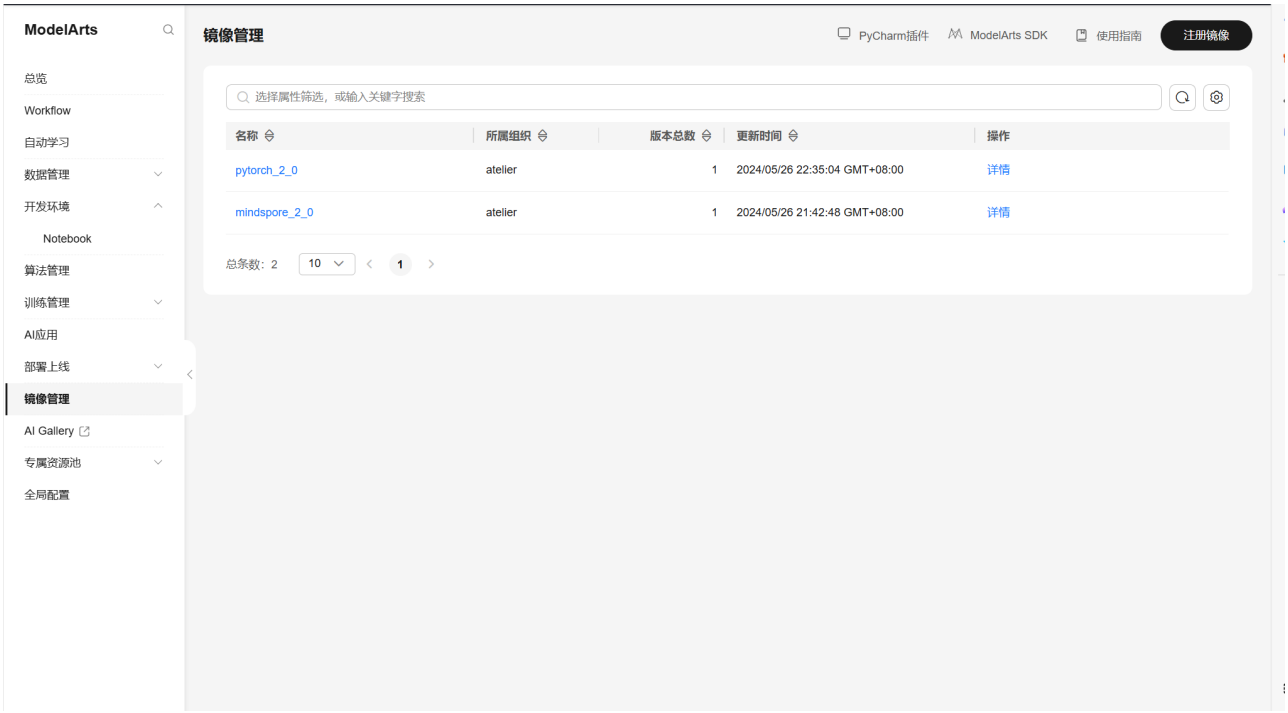
1. 找到并点击`前往AI Gallery获取更多镜像`

2. 点击第一条

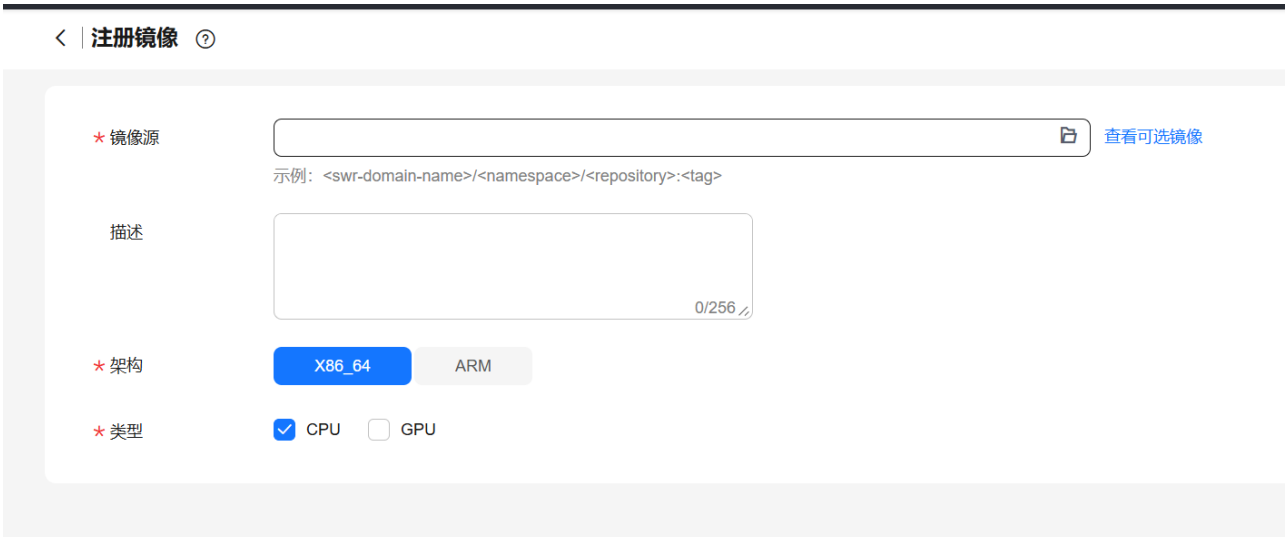


3. 复制他的url

4. 回到 modelArts， 点击注册镜像



5. 把复制的url粘贴于此



6. 注册完成后， 新建notebook时选择自定义镜像即可