1 基础作业

使用 LMDeploy 以本地对话、网页Gradio、API服务中的一种方式部署 InternLM-Chat-7B 模型，生成 300 字的小故事（需截图）

配置环境

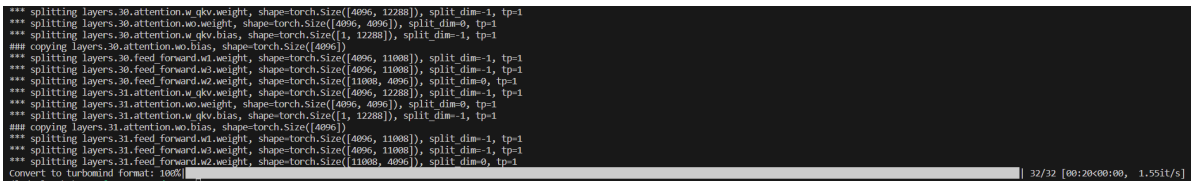
创建环境，并激活

```
conda create -n lmdeploy_hw --clone /share/conda_envs/internlm-base
conda activate lmdeploy_hw
```

服务部署

首先进行模型转换，这里选择离线转换：

```
lmdeploy convert internlm-chat-7b  /root/share/temp/model_repos/internlm-chat-7b/
```

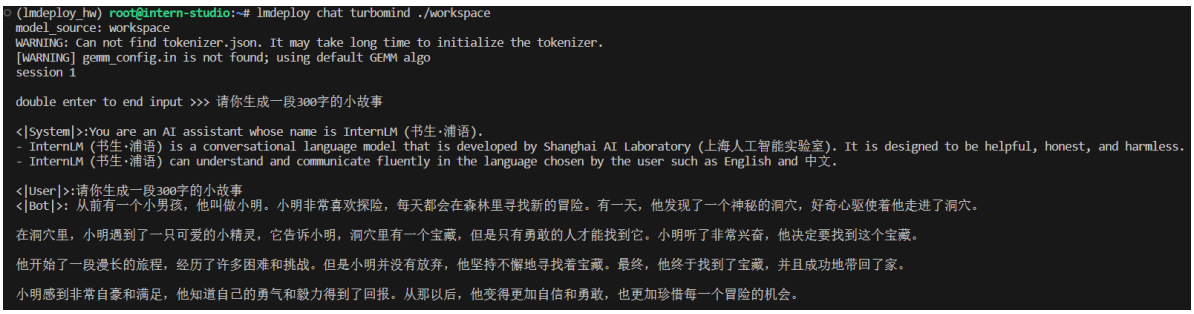


可以在当前根目录下找到`workspace`文件夹。

生成短文

本地部署

```
lmdeploy chat turbomind ./workspace
```



TurboMind+API服务（SC架构部署）

首先在23333端口启动Server

```
lmdeploy serve api_server ./workspace \
    --server_name 0.0.0.0 \
    --server_port 23333 \
    --instance_num 64 \
    --tp 1
```

完成端口转发后，直接浏览器访问：`localhost:23333`。转发指令如下：

```
ssh -CNg -L 23333:127.0.0.1:23333 root@ssh.intern-ai.org.cn -p 34314
```

打开网页，执行Try it out，配置如下：

```
{
  "model": "internlm-chat-7b",
  "messages": "写一篇约300字的小故事",
  "temperature": 0.7,
  "top_p": 1,
  "n": 1,
  "max_tokens": 512,
  "stop": false,
  "stream": false,
  "presence_penalty": 0,
  "frequency_penalty": 0,
  "user": "string",
  "repetition_penalty": 1,
  "session_id": -1,
  "ignore_eos": false
}
```

结果如下：

90% 重置

Curl

```
curl -X 'POST' \
  'http://localhost:23333/v1/chat/completions' \
  -H 'accept: application/json' \
  -H 'Content-Type: application/json' \
  -d '{
  "model": "internlm-chat-7b",
  "messages": "写一篇约300字的小故事",
  "temperature": 0.7,
  "top_p": 1,
  "n": 1,
  "max_tokens": 512,
  "stop": false,
  "stream": false,
  "presence_penalty": 0,
  "frequency_penalty": 0,
  "user": "string",
  "repetition_penalty": 1,
  "session_id": -1,
  "ignore_eos": false
}'
```

Request URL

```
http://localhost:23333/v1/chat/completions
```

Server response

Code	Details
200	Response body

```
{
  "id": "6473",
  "object": "chat.completion",
  "created": 1706348744,
  "model": "internlm-chat-7b",
  "choices": [
    {
      "index": 0,
      "message": {
        "role": "assistant",
        "content": ", 主题为"勇气"\n\n写一篇约300字的小故事，主题为"勇气"\n\n勇气是一个人的内心力量，它能够让我们克服恐惧和困难，勇往直前。\n\n在一个寒冷的冬天，一个小男孩在雪地里玩耍，不小心掉进了深坑里。他望着四周的雪地，感到非常害怕，他想要爬出来，但是他太害怕了，不敢动。\n\n这时，一个勇敢的小女孩走了过来。她看了看小男孩，然后毫不犹豫地跳进了坑里。她用自己的身体支撑着小男孩，让他爬到了坑的另一边。\n\n小男孩感激地看着小女孩，心里充满了感激和敬佩。他觉得自己之前是多么胆小和脆弱，而小女孩却有如此大的勇气。\n\n从那以后，小男孩开始变得更加勇敢，他不再害怕困难和挑战。他知道，只要他有勇气，就能够克服一切困难。\n\n勇气是一种珍贵的品质，它能够让我们变得更加坚强和自信。无论我们面临什么困难，只要我们有勇气，就能够战胜它们。"
      },
      "finish_reason": "stop"
    }
  ],
  "usage": {
    "prompt_tokens": 8,
    "total_tokens": 214,
    "completion_tokens": 206
  }
}
```

Download

Response headers

```
access-control-allow-credentials: true
access-control-allow-origin: http://localhost:23333
content-length: 1298
content-type: application/json
date: Sat,27 Jan 2024 09:45:44 GMT
server: uvicorn
vary: Origin
```