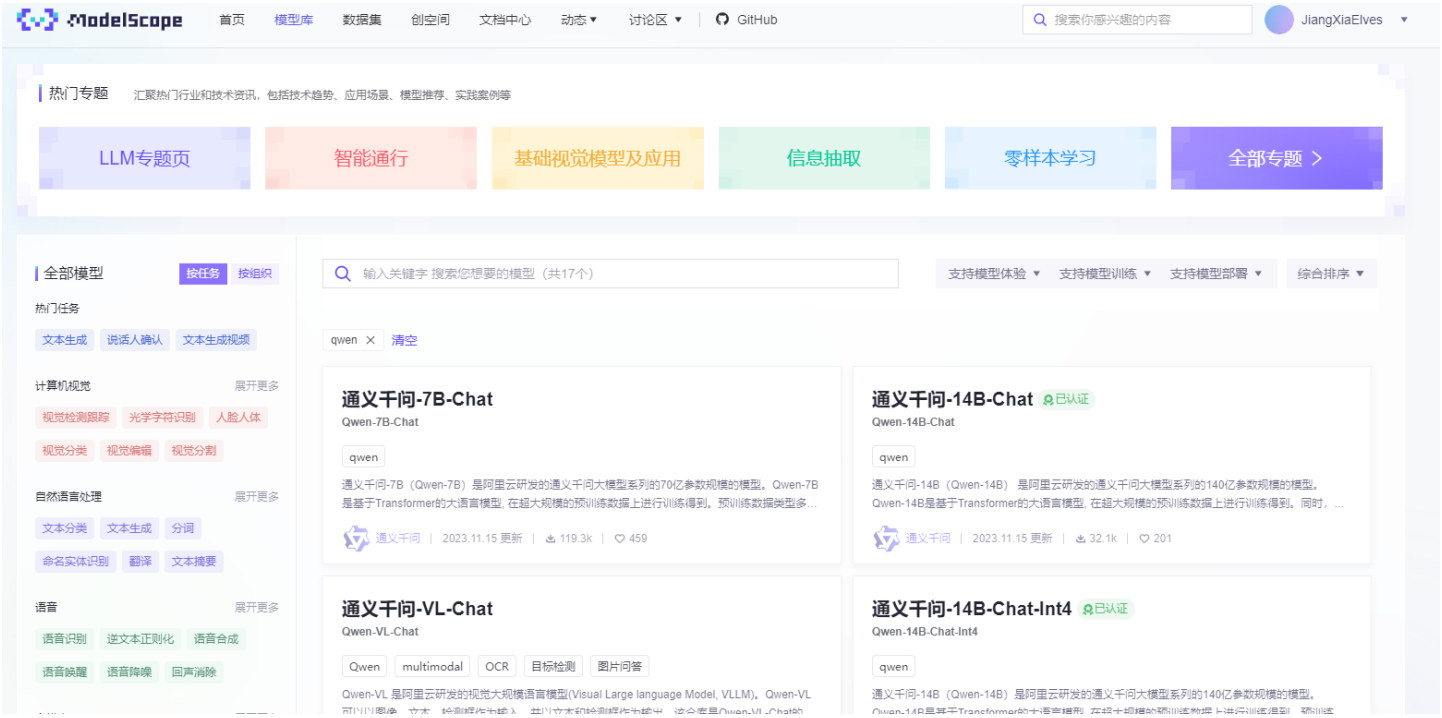
基于Qwen-14B-Chat的本地Agent—环境搭建

1. Qwen-14B 模型下载

在云服务器中，使用 Modelscope 平台的 modelscope 包进行模型下载是一种非常高效的方法。



1.1 创建 modelscope_env

```
conda create -n modelscope_env python=3.8.5
```

1.2 启动 modelscope_env

```
source activate modelscope_env
```

1.3 安装 modelscope 包

```
pip install modelscope
```

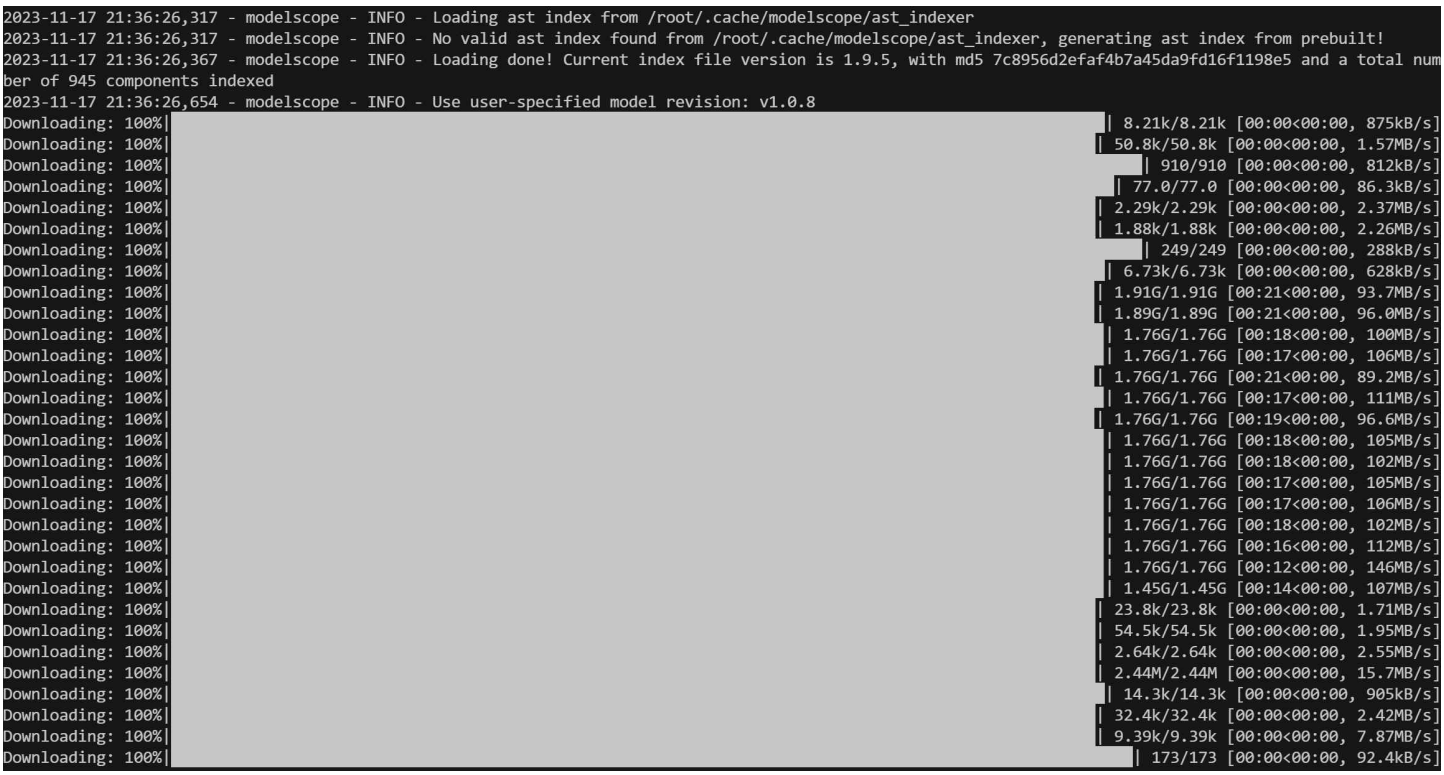
1.4 修改模型下载文件

将下载文件中的目录修改为你自己服务器的目录：

```
from modelscope.hub.snapshot_download import snapshot_download
model_dir = snapshot_download('qwen/Qwen-14B-Chat', cache_dir='你自己的目录',
revision='v1.0.8')
```

1.5 执行模型下载文件

```
python qwen_14B_chat_download.py
```



2. FastChat环境配置与启动

2.1 创建 fastchat_env

```
conda create -n fastchat_env python=3.8.5
```

2.2 启动 fastchat_env

```
source activate fastchat_env
```

2.3 安装FastChat包

```
bash fastchat_install.sh
```

2.4 安装 Qwen 依赖

安装 01_qwen_server 中 requirements.txt 对应的依赖：

```
pip install -r requirements.txt
```

2.5 启动 Controller

Controller 启动后会占用21001端口：

```
nohup bash controller_start.sh > cl_20231118.log &
```

```
(qwen_agent_env) root@autodl-container-faba4ab153-941bc3b5:~/autodl-tmp/jiangxia/01_qwen_server# lsof -i:21001
COMMAND  PID USER   FD   TYPE  DEVICE SIZE/OFF NODE NAME
python  2659 root    8u  IPv4 6816478      0t0  TCP *:21001 (LISTEN)
```

2.6 修改 Worker 目录

将worker_start.sh中的模型目录修改为你自己的目录：

```
python -m fastchat.serve.model_worker --model-path 你的模型目录 --host 0.0.0.0
```

2.7 启动 Worker

Controller 启动后会占用21002端口：

```
nohup bash worker_start.sh > wk_20231118.log &
```

```
(qwen_agent_env) root@autodl-container-faba4ab153-941bc3b5:~/autodl-tmp/jiangxia/01_qwen_server# lsof -i:21002
COMMAND  PID USER   FD   TYPE  DEVICE SIZE/OFF NODE NAME
python  3157 root   38u  IPv4 6824581      0t0  TCP *:21002 (LISTEN)
```

2.8 启动 OpenAI 接口支持

OpenAI 接口支持服务启动后会占用8000端口：

```
nohup bash openai_support.sh > oa_20231118.log &
```

```
(qwen_agent_env) root@autodl-container-faba4ab153-941bc3b5:~/autodl-tmp/jiangxia/01_qwen_server# lsof -i:8000
COMMAND  PID USER   FD   TYPE  DEVICE SIZE/OFF NODE NAME
python  3219 root    7u  IPv4 6761284      0t0  TCP *:8000 (LISTEN)
```

2.9 在本地机器建立端口映射

```
ssh -CNg -L 8000:127.0.0.1:8000 root@region-3.seetacloud.com -p 47819
```