

# HUAWEI体验课

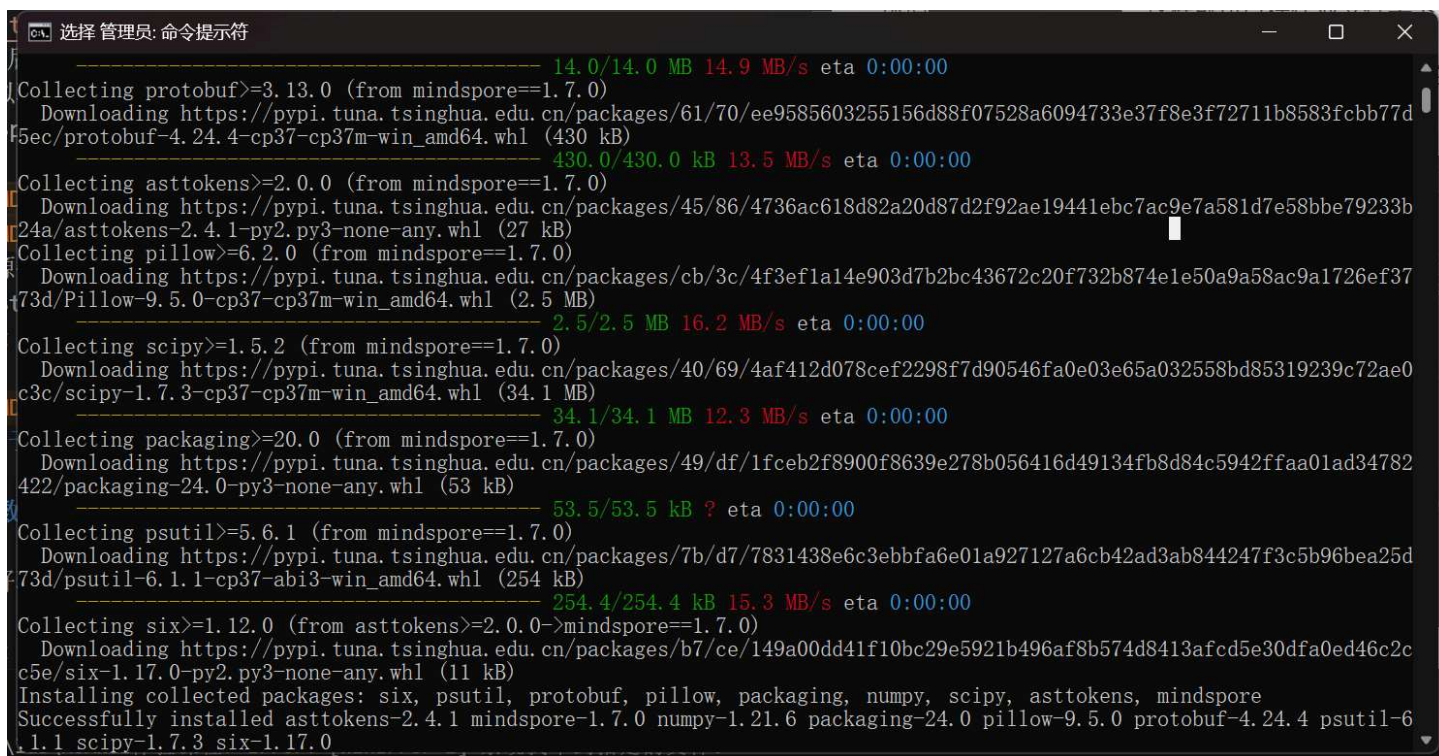
2211771 原敬闰

## 环境配置

首先需要安装 python 3.7.5，但是我的电脑已经安装了 3.11.4，所以陷入僵局，执行命令：

```
python -m pip install https://ms-release.obs.cn-north-4.myhuaweicloud.com/1.7.0/MindSpore/cpu/;
```

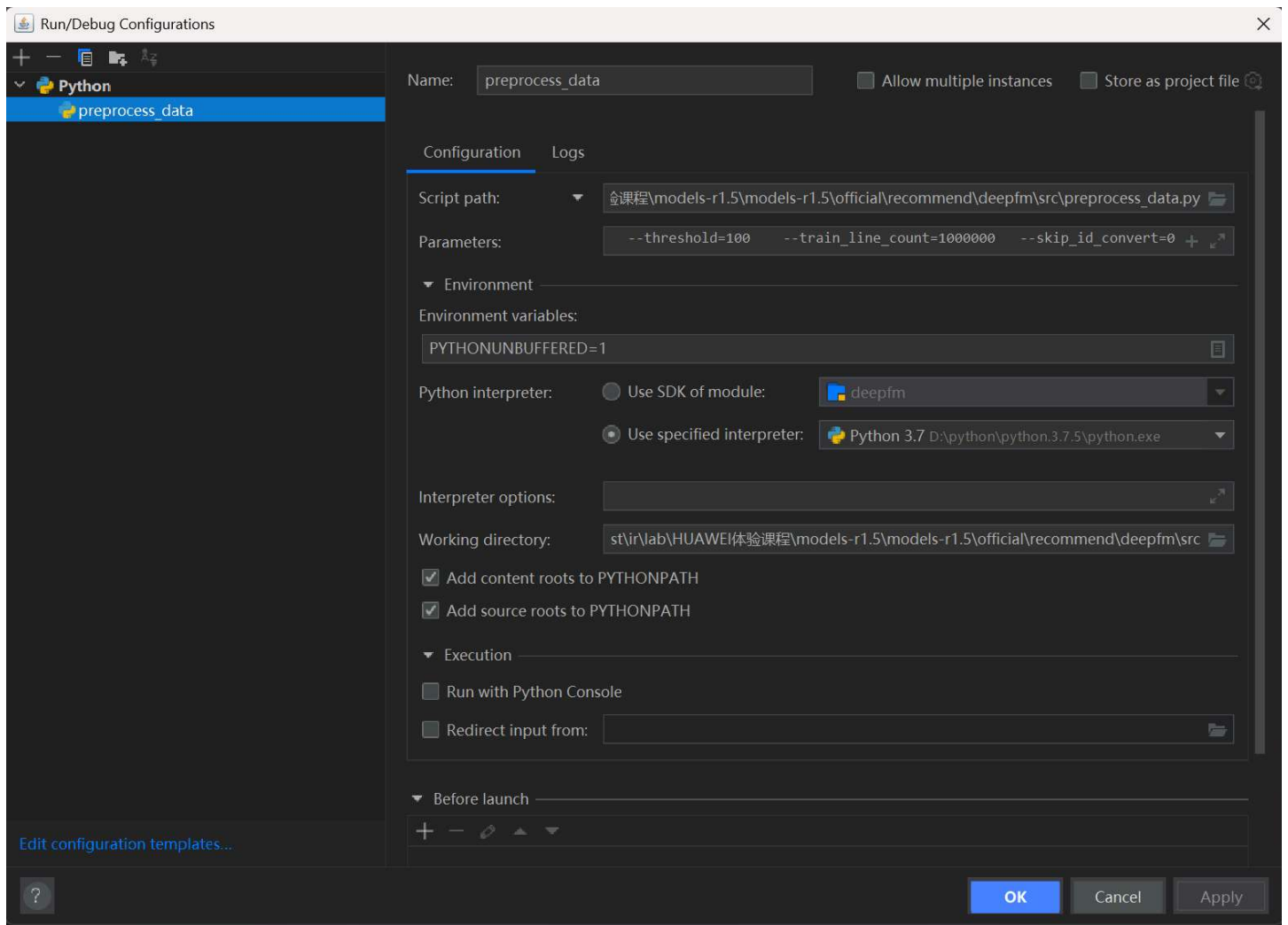
的时候一直在报错，最后只好下载了 3.7.5 之后暂时将 3.11 中的可执行程序改名，随后在作业文件夹下配置一个 3.7 的虚拟环境然后将名字改回来，这样就可以在作业文件夹下下载 MindSpore。



```
选择 管理员: 命令提示符
Collecting protobuf>=3.13.0 (from mindspore==1.7.0)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/61/70/ee9585603255156d88f07528a6094733e37f8e3f72711b8583fcbb77d5ec/protobuf-4.24.4-cp37-cp37m-win_amd64.whl (430 kB)
  14.0/14.0 MB 14.9 MB/s eta 0:00:00
Collecting asttokens>=2.0.0 (from mindspore==1.7.0)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/45/86/4736ac618d82a20d87d2f92ae19441ebc7ac9e7a581d7e58bbe79233b24a/asttokens-2.4.1-py2.py3-none-any.whl (27 kB)
  430.0/430.0 kB 13.5 MB/s eta 0:00:00
Collecting pillow>=6.2.0 (from mindspore==1.7.0)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/cb/3c/4f3ef1a14e903d7b2bc43672c20f732b874e1e50a9a58ac9a1726ef37473d/Pillow-9.5.0-cp37-cp37m-win_amd64.whl (2.5 MB)
  2.5/2.5 MB 16.2 MB/s eta 0:00:00
Collecting scipy>=1.5.2 (from mindspore==1.7.0)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/40/69/4af412d078cef2298f7d90546fa0e03e65a032558bd85319239c72ae0c3c/scipy-1.7.3-cp37-cp37m-win_amd64.whl (34.1 MB)
  34.1/34.1 MB 12.3 MB/s eta 0:00:00
Collecting packaging>=20.0 (from mindspore==1.7.0)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/49/df/1fceb2f8900f8639e278b056416d49134fb8d84c5942ffaa01ad34782422/packaging-24.0-py3-none-any.whl (53 kB)
  53.5/53.5 kB ? eta 0:00:00
Collecting psutil>=5.6.1 (from mindspore==1.7.0)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/7b/d7/7831438e6c3ebbfa6e01a927127a6cb42ad3ab844247f3c5b96bea25d73d/psutil-6.1.1-cp37-abi3-win_amd64.whl (254 kB)
  254.4/254.4 kB 15.3 MB/s eta 0:00:00
Collecting six>=1.12.0 (from asttokens>=2.0.0->mindspore==1.7.0)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/b7/ce/149a00dd41f10bc29e5921b496af8b574d8413afcd5e30dfa0ed46c2cc5e/six-1.17.0-py2.py3-none-any.whl (11 kB)
Installing collected packages: six, psutil, protobuf, pillow, packaging, numpy, scipy, asttokens, mindspore
Successfully installed asttokens-2.4.1 mindspore-1.7.0 numpy-1.21.6 packaging-24.0 pillow-9.5.0 protobuf-4.24.4 psutil-6.1.1 scipy-1.7.3 six-1.17.0
```

## 基于 MindStudio 的 DeepFM 点击率预估实验

下载源代码和数据并使用 MindStudio 打开，随后按照实验指导书下载 MindStudio 并为项目配置 python 编译环境：



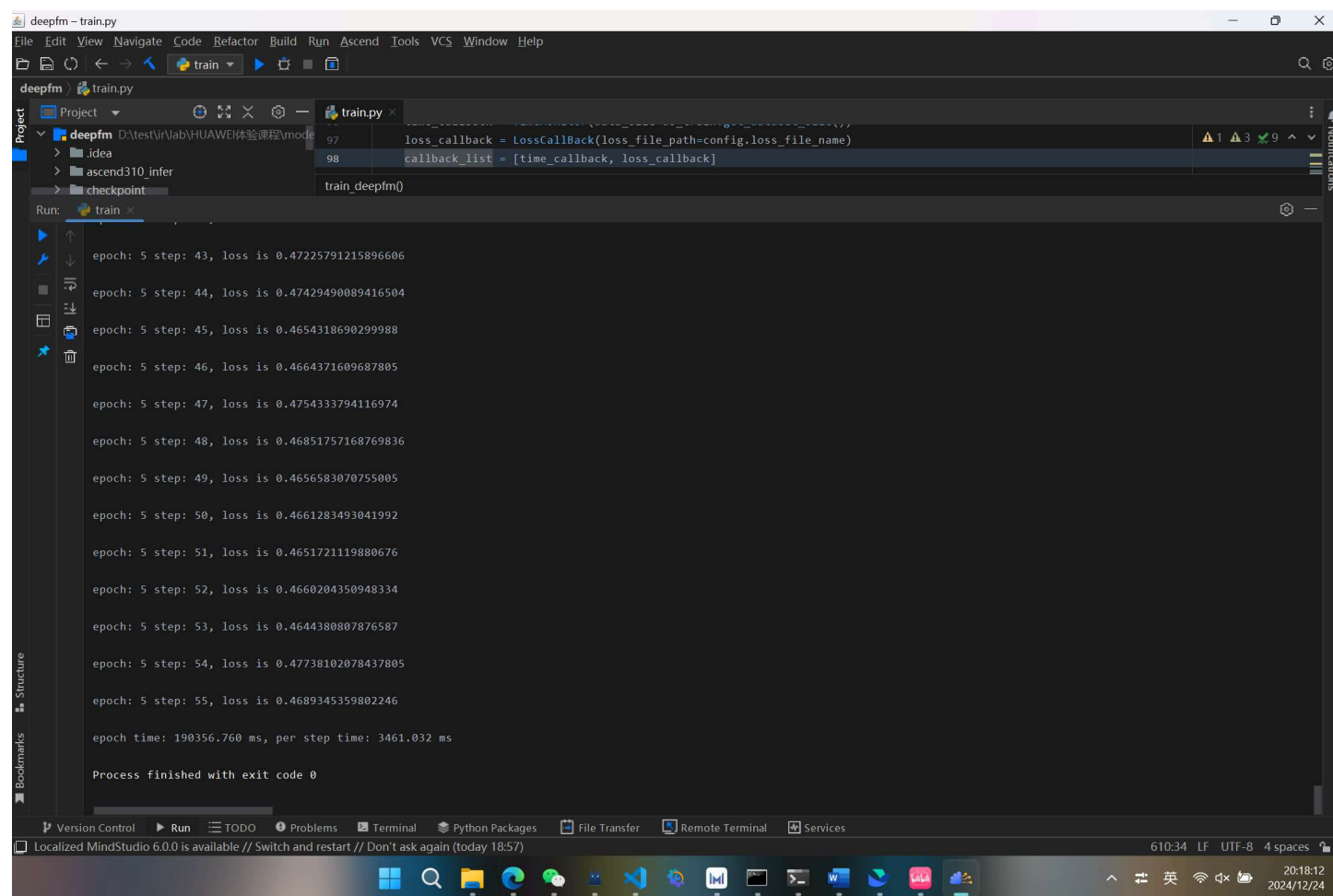
## 数据集预处理

运行 preprocess\_data.py :



# 模型训练

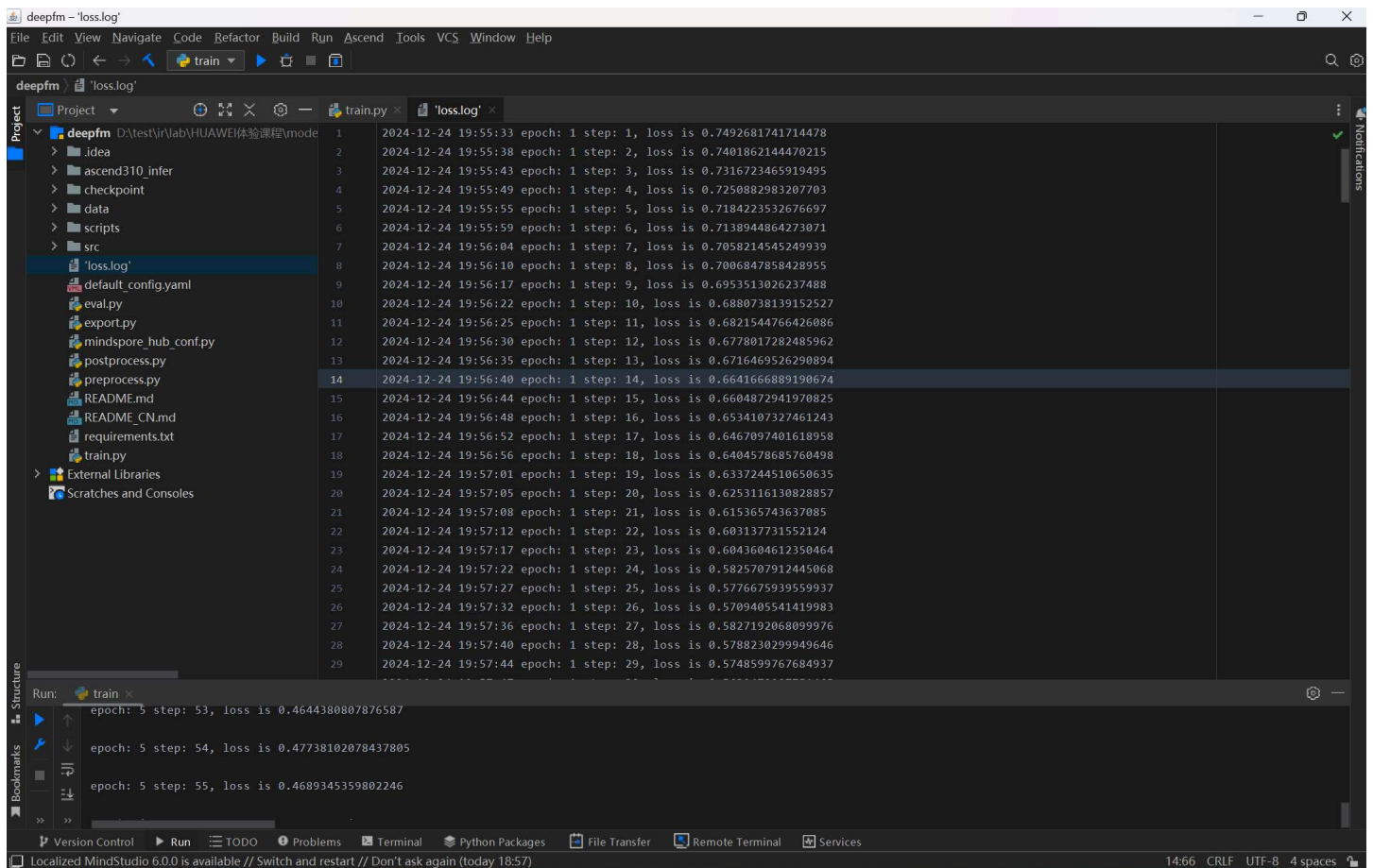
运行 `train.py` 文件进行模型训练，可以看到训练过程中输出的损失值：



```
deepfm - train.py
File Edit View Navigate Code Refactor Build Run Ascend Tools VCS Window Help
deepfm \train.py
Project
  deepfm D:\test\lab\HUAWEI\体验课程\mode
    .idea
    ascend310_infer
    checkpoint
Run: train
epoch: 5 step: 43, loss is 0.47225791215896606
epoch: 5 step: 44, loss is 0.47429490089416504
epoch: 5 step: 45, loss is 0.4654318690299988
epoch: 5 step: 46, loss is 0.4664371609687805
epoch: 5 step: 47, loss is 0.4754333794116974
epoch: 5 step: 48, loss is 0.46851757168769836
epoch: 5 step: 49, loss is 0.4656583070755005
epoch: 5 step: 50, loss is 0.4661283493041992
epoch: 5 step: 51, loss is 0.4651721119880676
epoch: 5 step: 52, loss is 0.4660204350948334
epoch: 5 step: 53, loss is 0.4644380807876587
epoch: 5 step: 54, loss is 0.47738102078437805
epoch: 5 step: 55, loss is 0.4689345359802246
epoch time: 190356.760 ms, per step time: 3461.032 ms
Process finished with exit code 0
```

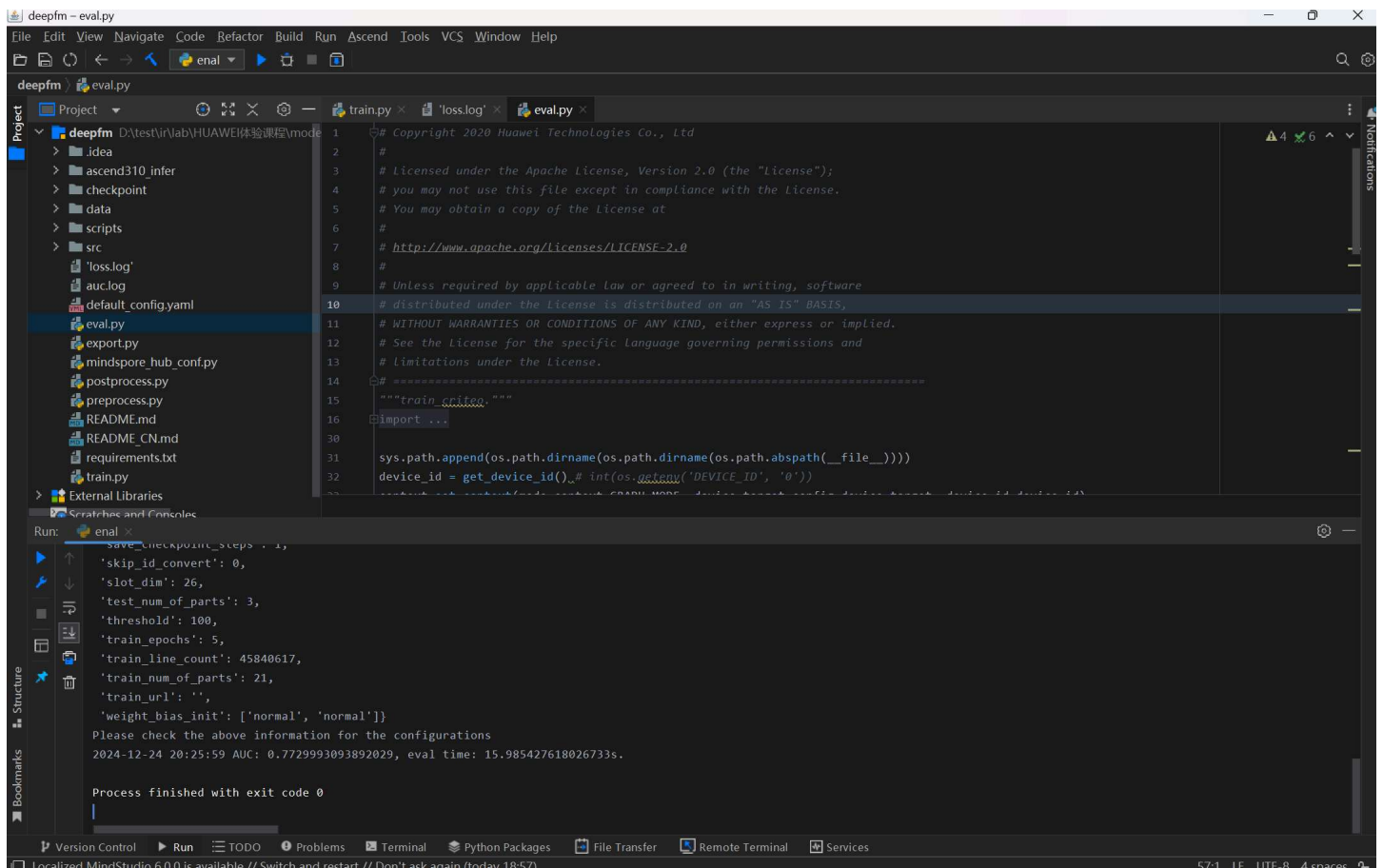
也可以在 `loss.log` 文件夹中查看项目损失值：



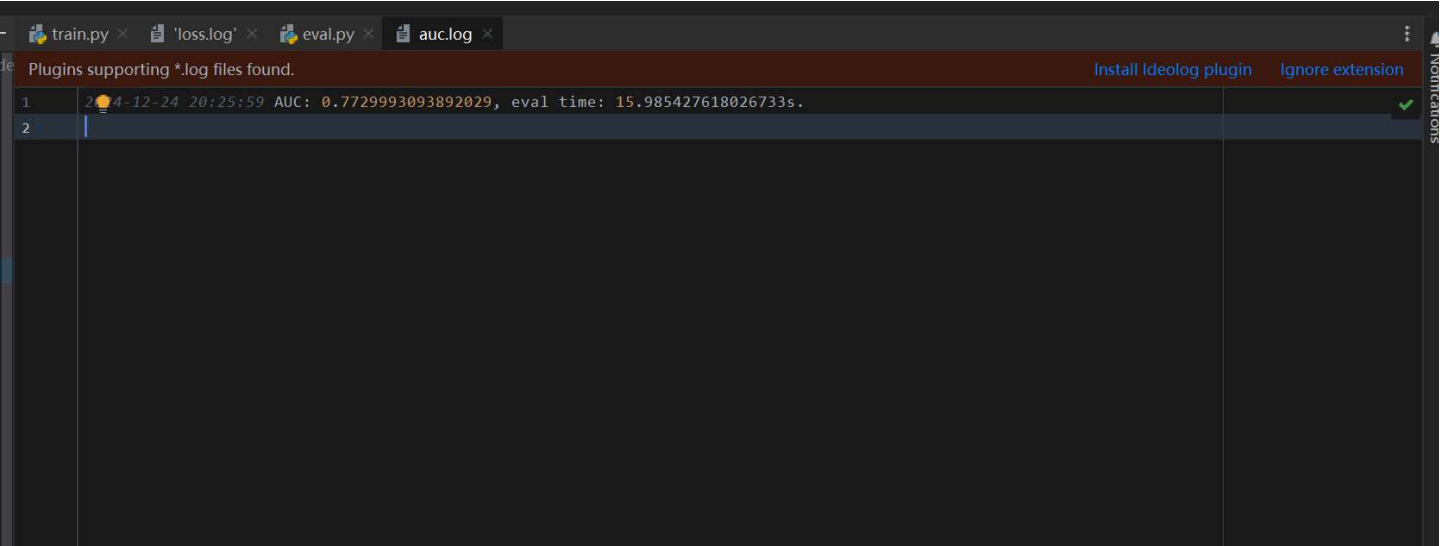


## 模型评估

运行 eval.py 文件进行评估:

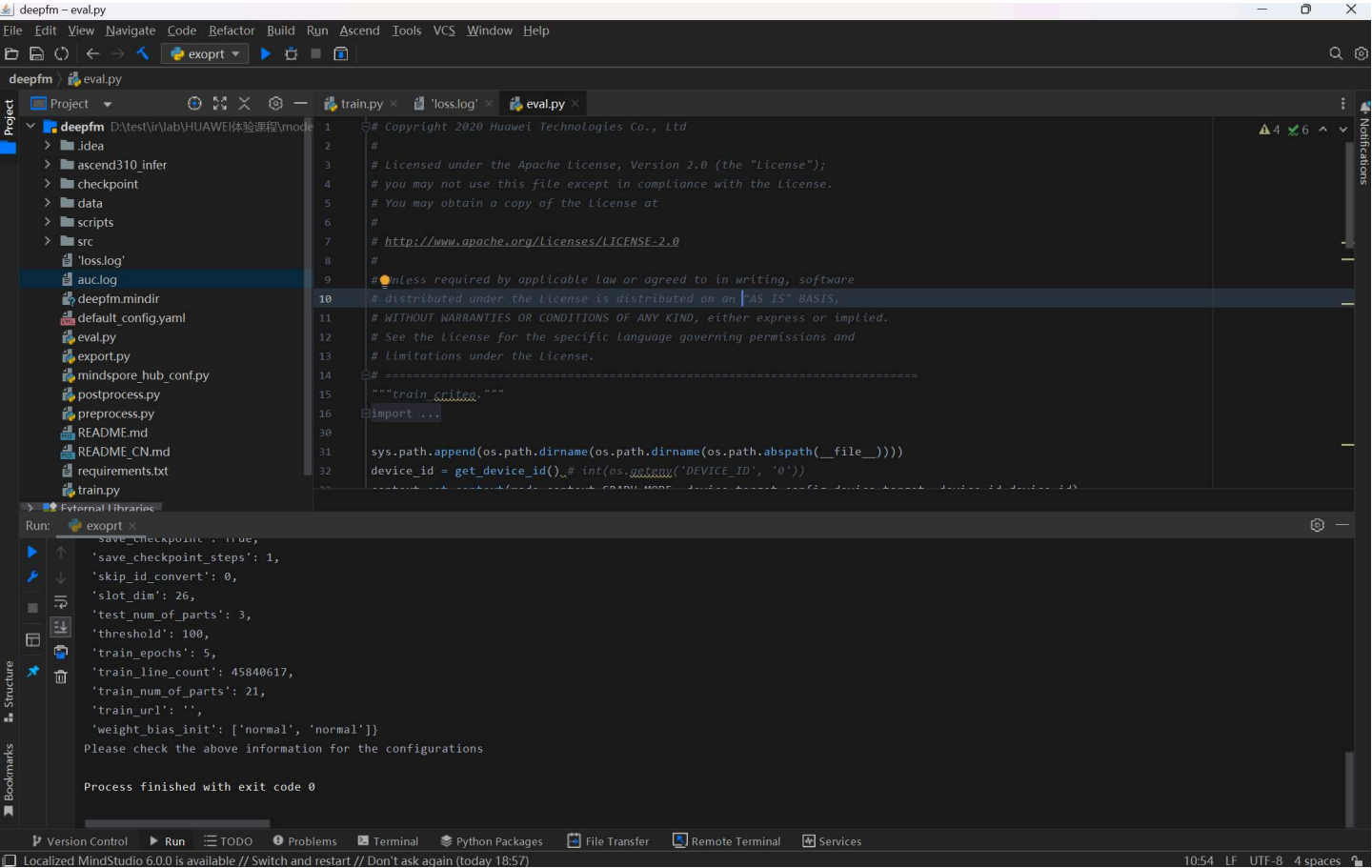


随后在 auc.log 文件中查看准确率日志：

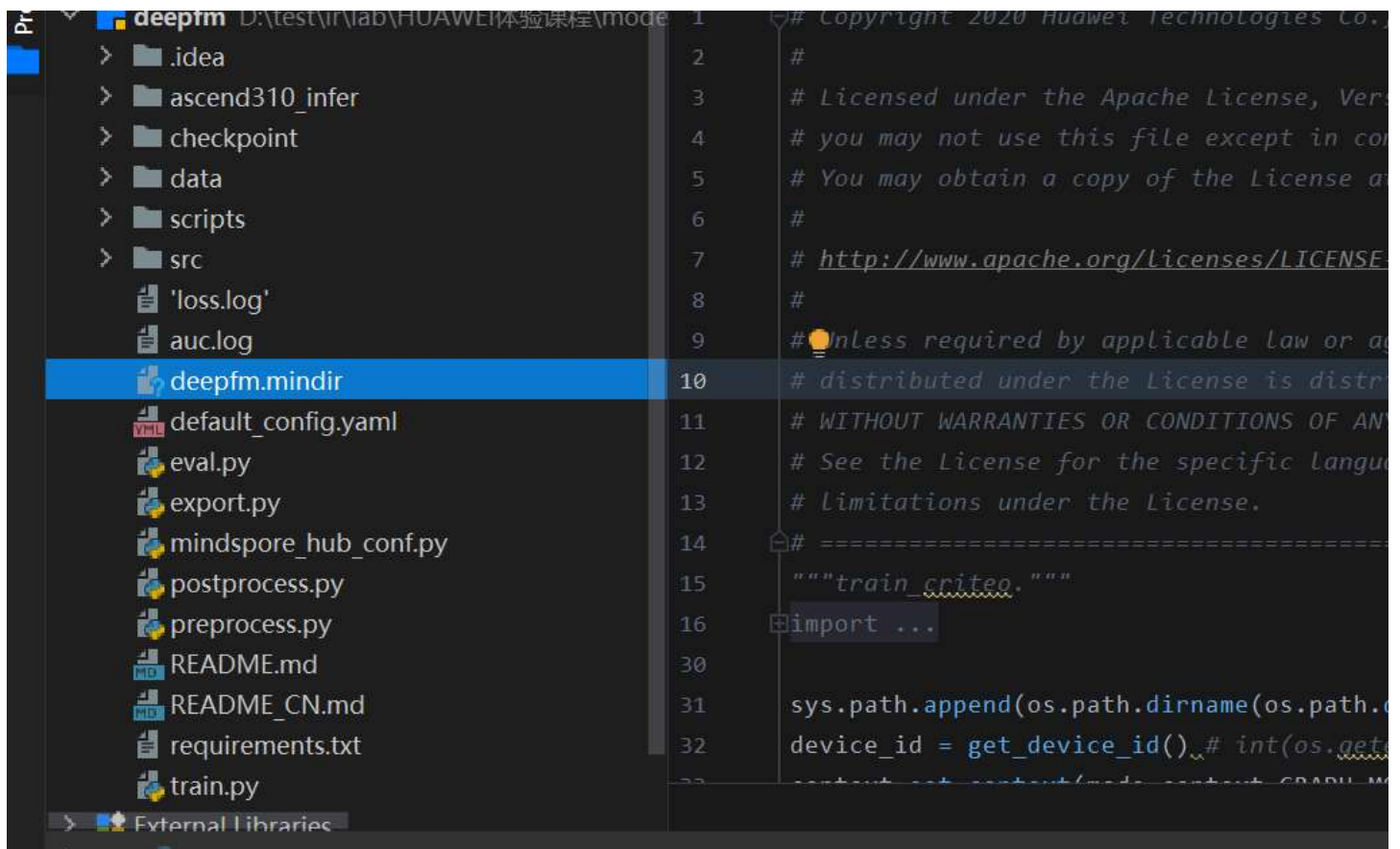


## 导出 MindIR

执行 export.py 脚本以导出模型文件：



可以看到目录下的 deepfm.mindir 模型文件：



## 实验中遇到的问题

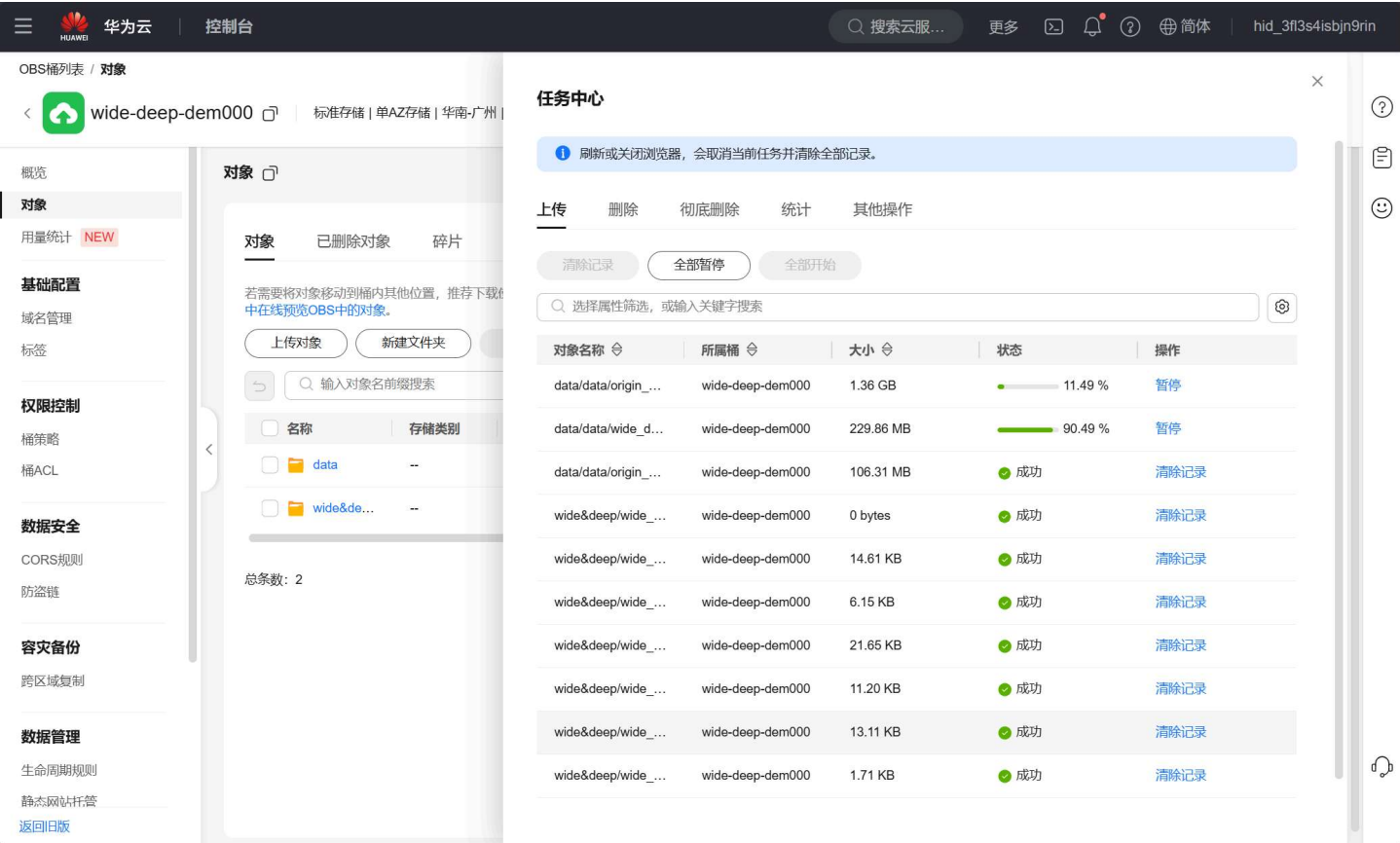
遇到的主要问题是配置环境的问题，上面已经叙述过，其次的问题在于第一次运行完 preprocess\_data.py 后，执行 train.py 莫名其妙的出错：显示某个文件不存在或没有权限。

于是我尝试重新运行 preprocess\_data.py 文件，出错了，于是删除掉所有第一次的输出，预处理又可以运行了，所以重新运行 train.py，然后就成功解决了这个问题，虽然我也不知道为什么第一次运行会出错。

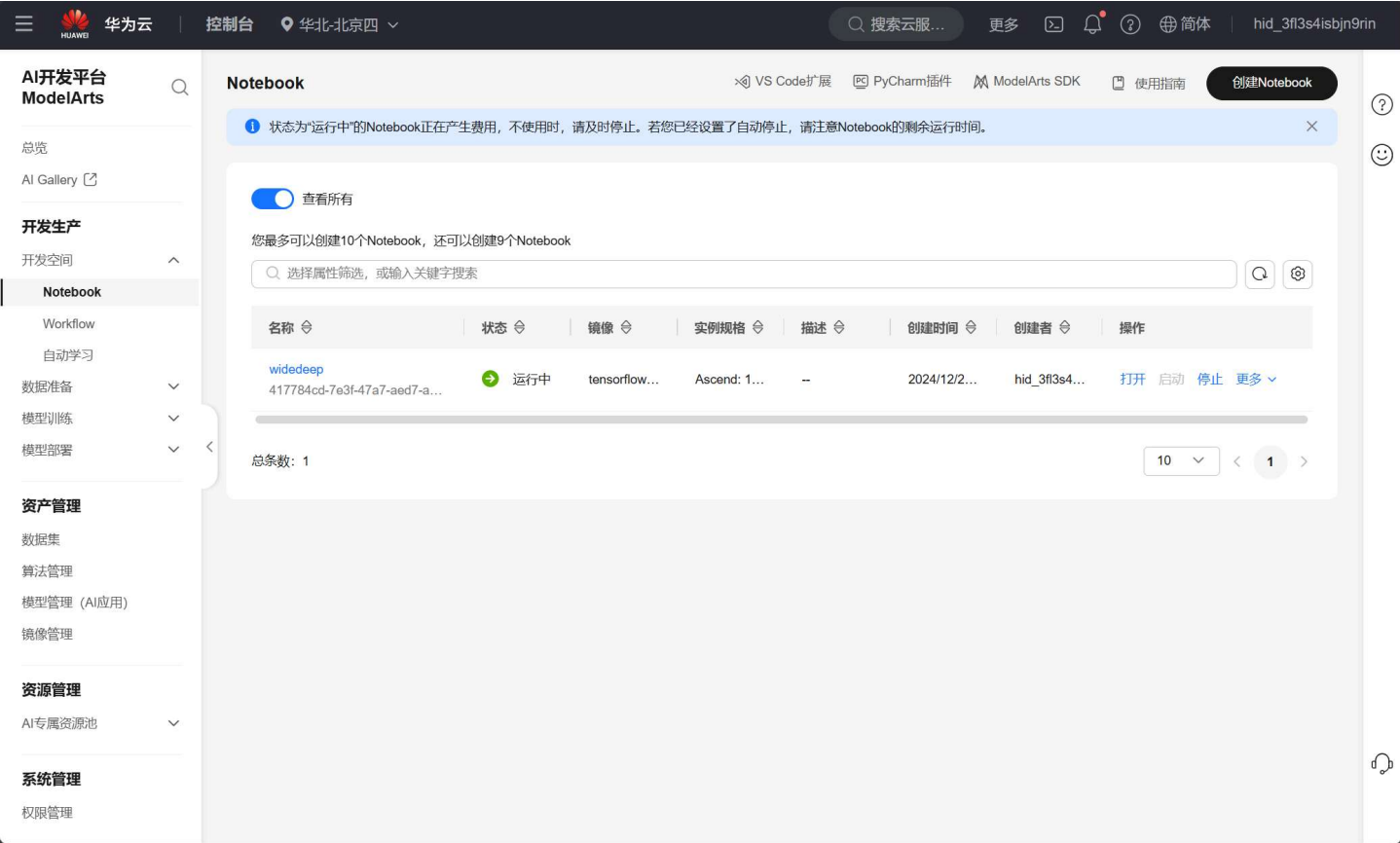
## 基于 MindSpore 的广告推荐

### 环境配置

首先在 OBS 新建一个桶，并将所需要的源码和数据集上传：



在华为云控制台新建一个 Notebook：



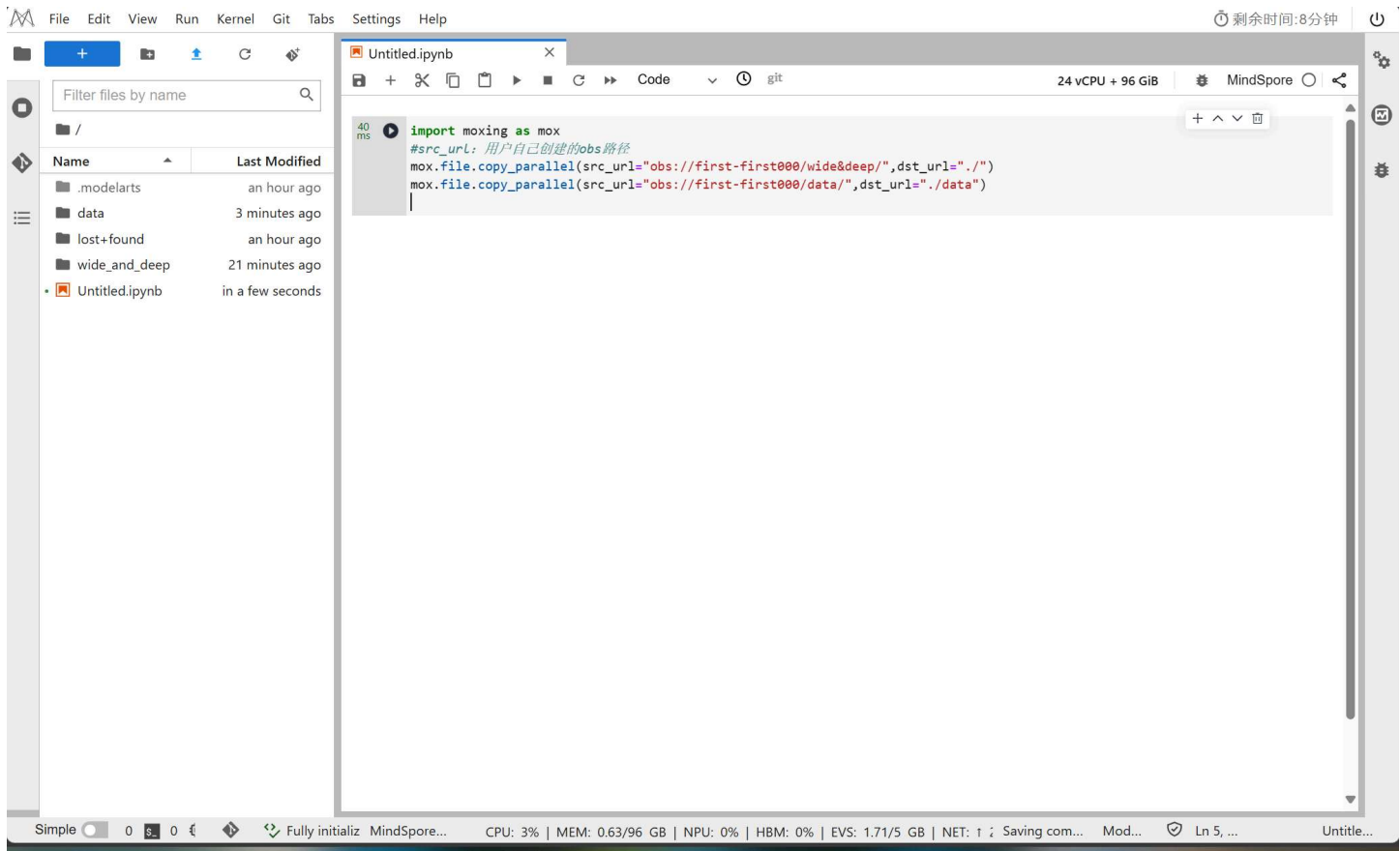
运行代码将桶中的数据和源码导入进来：



```
import moxing as mox

#src_url: 用户自己创建的obs路径

mox.file.copy_parallel(src_url="obs://first-first000/wide&deep/",dst_url="./")
mox.file.copy_parallel(src_url="obs://first-first000/data/",dst_url="./data")
```



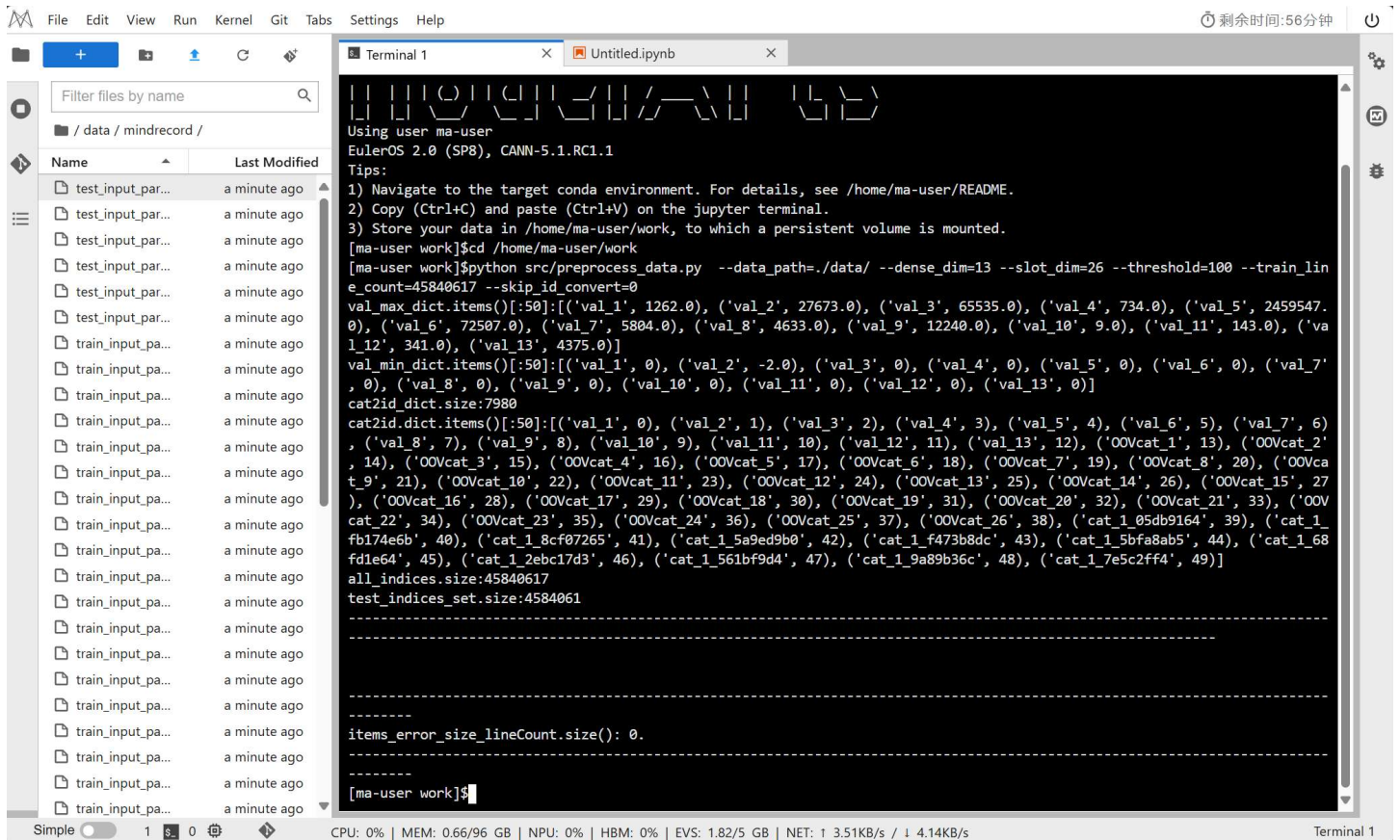
## 数据处理

进入到 /home/ma-user/work 路径下，在终端中执行命令：

```
python src/preprocess_data.py --data_path=./data/ --dense_dim=13 --slot_dim=26 --threshold=100
```



可以看到 data 目录下生成了 mindrecord 文件夹，里面存放了 mindrecord 数据，结果如图：



## 模型训练

进入到 `/home/ma-user/work` 路径下，在终端中执行命令：

```
python train_and_eval.py --data_path=./data/mindrecord --data_type=mindrecord
```

可以看到如下的训练过程输出信息：

File Edit View Run Kernel Git Tabs Settings Help

Filter files by name

Name	Last Modified
rank_0	5 minutes ago
script	an hour ago
src	5 minutes ago
eval.log	4 minutes ago
eval.py	an hour ago
export.py	an hour ago
items_error_size_lineCount.npy	17 minutes ago
loss.log	4 minutes ago
mindspore_hub_conf.py	an hour ago
README_CN.md	an hour ago
README.md	an hour ago
requirements.txt	an hour ago
train_and_eval_auto_parallel.py	an hour ago
train_and_eval_distribute.py	an hour ago
train_and_eval_parameter_server_distribute.py	an hour ago
train_and_eval_parameter_server_standalone.py	an hour ago
train_and_eval.py	an hour ago
train.py	an hour ago
Untitled.ipynb	26 minutes ago
widedeep_train-11_25.ckpt	4 minutes ago
widedeep_train-12_25.ckpt	4 minutes ago
widedeep_train-13_25.ckpt	4 minutes ago
widedeep_train-14_25.ckpt	4 minutes ago
widedeep_train-15_25.ckpt	4 minutes ago
widedeep_train-graph.meta	4 minutes ago

Simple 1 0

CPU: 0% | MEM: 0.66/96 GB | NPU: 0% | HBM: 0% | EVS: 2.95/5 GB | NET: ↑ 3.50KB/s / ↓ 4.11KB/s

Terminal 1

```
6200388
2024-12-25 16:0403 == Rank: 0 == EvalCallBack model.eval(): dict_values([0.7590292526200388]); eval_time: 0s
===loss=== 0 12 25 0.47633076 0.47820264
epoch: 12, step: 25, wide_loss: 0.47633076, deep_loss: 0.47820264
epoch time: 1358.471 ms, per step time: 54.339 ms
===== auc_metric end
===== auc: 0.759470088
9081107
2024-12-25 16:0404 == Rank: 0 == EvalCallBack model.eval(): dict_values([0.7594700889081107]); eval_time: 0s
===loss=== 0 13 25 0.46542916 0.46737117
epoch: 13, step: 25, wide_loss: 0.46542916, deep_loss: 0.46737117
epoch time: 1332.493 ms, per step time: 53.300 ms
===== auc_metric end
===== auc: 0.759851368
0457645
2024-12-25 16:0405 == Rank: 0 == EvalCallBack model.eval(): dict_values([0.7598513680457645]); eval_time: 0s
===loss=== 0 14 25 0.46409002 0.46612325
epoch: 14, step: 25, wide_loss: 0.46409002, deep_loss: 0.46612325
epoch time: 1387.080 ms, per step time: 55.483 ms
===== auc_metric end
===== auc: 0.760065783
349015
2024-12-25 16:0407 == Rank: 0 == EvalCallBack model.eval(): dict_values([0.76006578349015]); eval_time: 0s
===loss=== 0 15 25 0.46761012 0.46974477
epoch: 15, step: 25, wide_loss: 0.46761012, deep_loss: 0.46974477
epoch time: 1529.560 ms, per step time: 61.182 ms
===== auc_metric end
===== auc: 0.759185860
3445574
2024-12-25 16:0409 == Rank: 0 == EvalCallBack model.eval(): dict_values([0.7591858603445574]); eval_time: 0s
[ma-user work]$
```

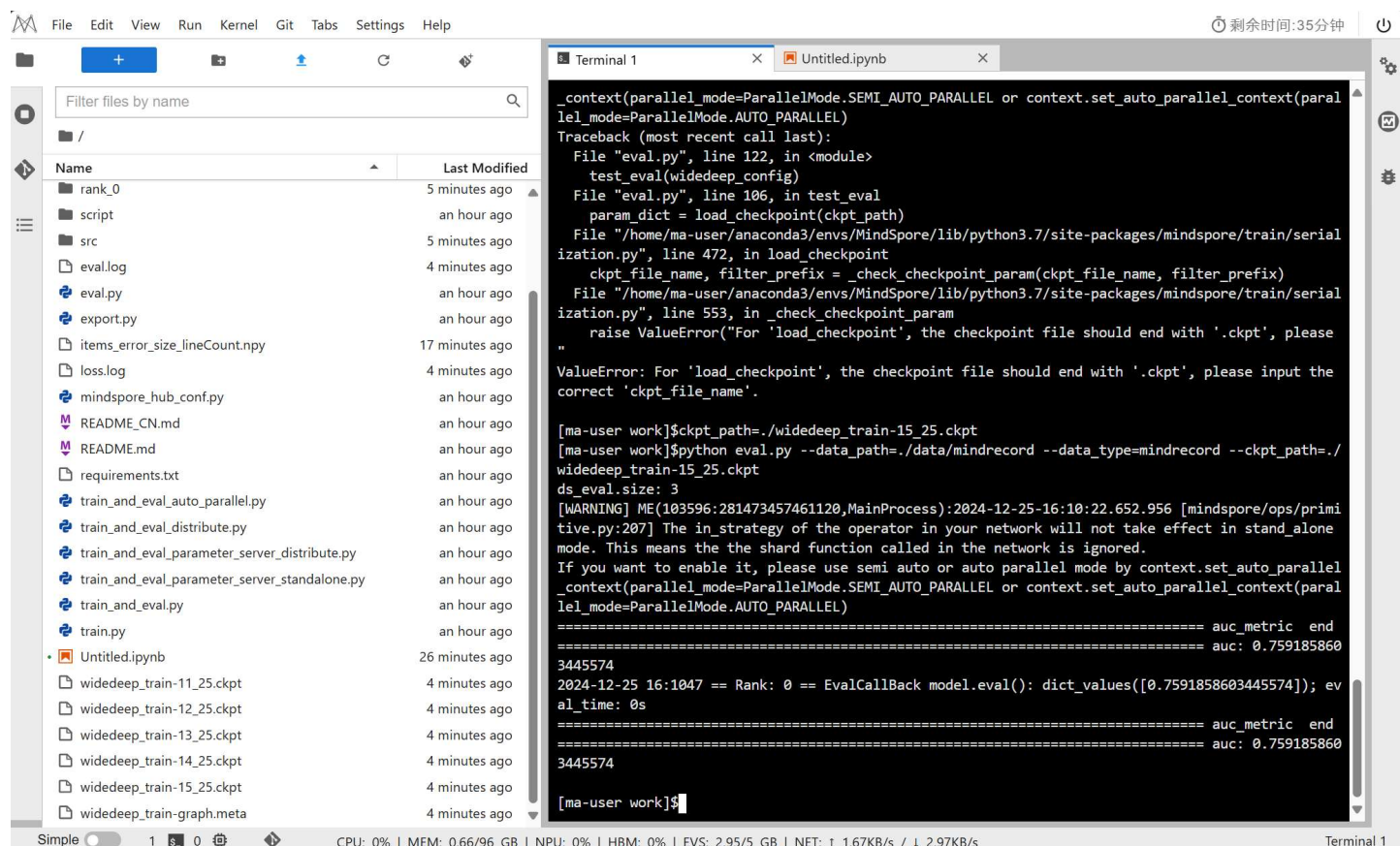
在左侧列表中可以看到生成的模型文件。

## 模型评估

进入到 `/home/ma-user/work` 路径下，在终端中执行命令：

```
python eval.py --data_path=./data/mindrecord --data_type=mindrecord --ckpt_path=./widedeep_train
```

模型评估结果如下：



## 遇到的问题

1. 首先是数据始终无法导入到 notebook 中，询问助教尝试修改权限等等之后仍然不行，最后助教提到要将二者放在同一个区中，于是重新创建一个桶，可以成功导入文件。
2. 上传源码和数据的时候一致失败，尤其是较大的文件夹，应该是网络问题，一直重试才成功。