先进入目录/blue/qsong1/wang.qing/benchmark\_scLLM\_API/UCE-main/UCE-main/

1.

先运行dataset\_making.py文件，将csv转为适合的输入。

由于每个模型的输入不太一样所以不同的csv的到输出可能不一样有些是pkl有些是ndarray。

这里主要要考虑的参数就是is\_sorted=True, seq\_length=8192

，别的就默认就行

一个是sequence是否按表达量排序，另一个是sequence的长度限制。

还有文件路径需要更改

2.

针对fixed embedding还要运行get\_ebd.py得到模型的embedding。

3.

针对fixed embedding运行benchmarking\_main\_EBD.py进行训练和预测

这里考虑的参数，别的就默认就行

parser.add\_argument('--train\_rate', type=float, default=0.8, help='')

控制训练和测试的比例

4.

针对fine tuning运行benchmarking\_main\_FT.py进行训练和预测

这里考虑的参数，别的就默认就行

parser.add\_argument('--ft\_list', type=list, default=['out\_proj'], help='')

控制lora微调的网络层。

parser.add\_argument('--train\_rate', type=float, default=0.8, help='')

控制训练和测试的比例。

注意保持目录的结构，更改可能会导致package冲突。

总共8个模型

现在UCE已经整理完成了，先按照UCE构建项目，其余的模型框架和UCE差不多的。

直接在hypegator复制会比较方便，模型数据也比较大。

代码在/blue/qsong1/wang.qing/benchmark\_scLLM\_API/UCE-main/

数据在/blue/qsong1/wang.qing/benchmark\_dataset\_API/UCE/