# 报告

基本在debug，并且没有成功debug结束

首先是关于gpu数量的问题，通过之前的debug发现可能是bitahub网站安装invidia的driver api的版本低于runtime 的api，查阅网上一些教程发现是可能会有问题，但后面发现是由于bash脚本中设定的gpu数量为8个，自己跑模型的时候用的是四个，所以出问题了。

在脚本路径方面，作者script拼写少了个t导致脚本路径出现了问题。在下载pretrain模型时，那个脚本的配置会显示json文件的的bug，通过debug发现是由于他路径设置有问题，查找源代码单独设定了—model\_name\_or\_path –congig\_name,--tokenizer\_name等分支，并且前两者要具体到相应的json文件，最后一个是token对应的文件夹就行了。

在使用hugging face提供的datasets.load\_dataset()函数的时候，出现了connection error。试过采用git clone进行下载，首先遇到的问题是load\_dataset函数的第二个变量config\_name,这个我并不知道是什么东西，注释给的是数据集的配置，但是在git clone中的-c分支的配置似乎并不相同。在尝试直接加入分支的时候提醒我需要输入git hub的账号和密码，这个我多次输入自己的github账号，邮箱和密码后都没有效果。最后决定舍弃这一分支。利用git clone成功下载hugging face上的xsum数据集的时候，发现他的数据集是lfs文件的一个网站，而不是大数据集本身，这个尝试在bithub网站上安装git lfs文件，但是好像是需要sudo权限才可以安装，经过多次其他尝试失败后决定直接用data\_loader加载lfs文件，在debug过程中出现了加载选项，但是一直出现了connection error，调试中分别出现了401 Unauthorized错误，和104:connetction reset by peer错误，前者可能是登录账号权限的问题，但是经过适当修改后这一错误基本解决。通过查阅CSDN和github社区发现后面的错误就是dataset里面的的load\_dataset函数自身连接存在着问题，并且即使在最新版本这一错误也并没有修复。因而最终数据集在多次尝试后加载失败。