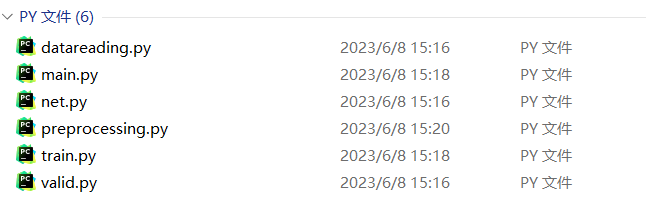
该项目下属5个.py文件以及数据预处理文件。

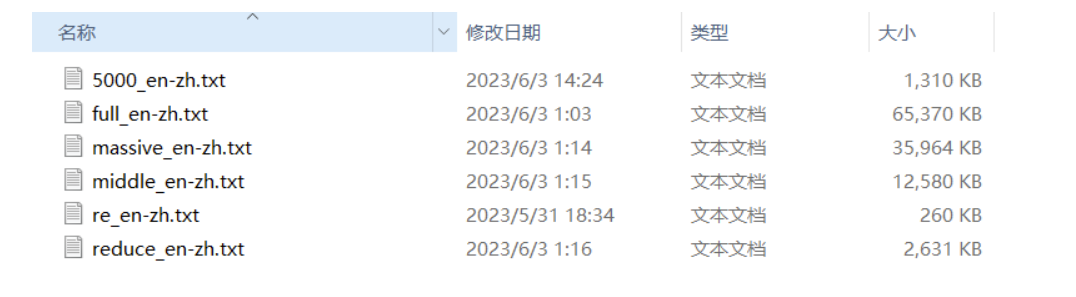


其中datareading.py为读取数据，net为定义网络，train为定义训练代码，valid定义验证代码，main是主程序，包含训练和验证。Preprocessing.py文件为数据预处理文件，在进行项目之前先对原数据进行预处理。

在进行训练时，运行main代码，其会调用train代码和val代码。而train和valid会调用net以及数据预处理代码。

所需环境：

pytorch，jieba，matplotlib，nltk



massive\_en-zh.txt包含15w数据

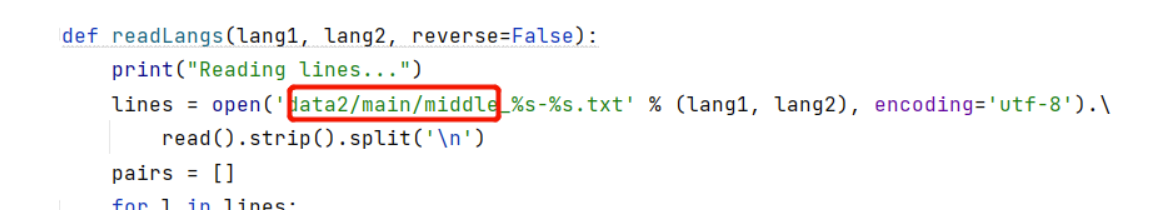
middle\_en-zh.txt包含5w数据

reduce\_en-zh.txt包含1w数据

5000\_en-zh.txt包含5000数据

re\_en-zh.txt包含1000数据，用于测试

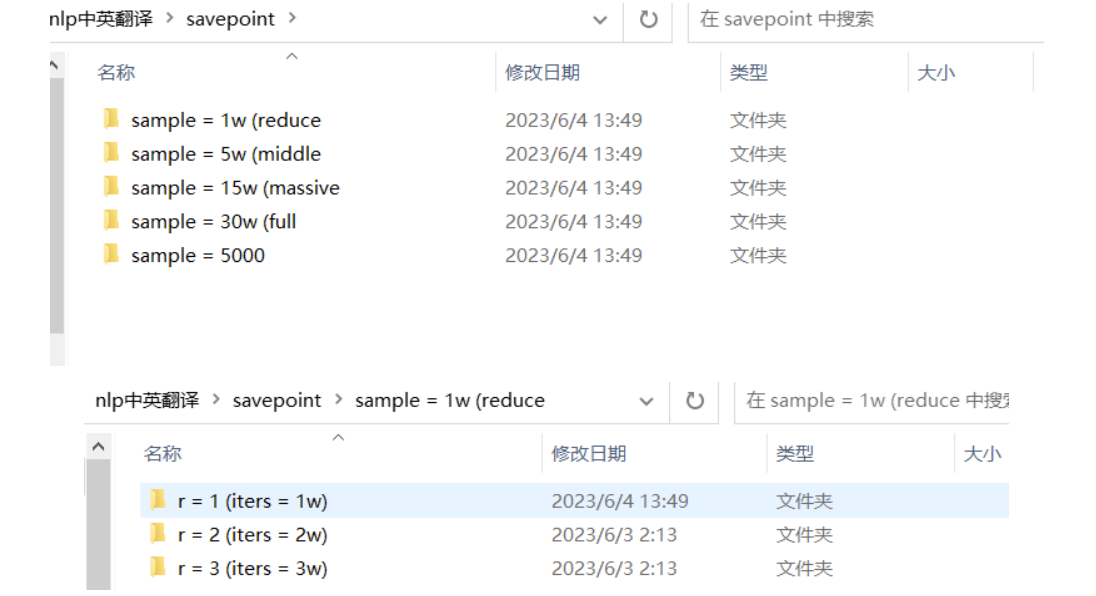
修改导入文件的地址只需要修改前缀即可



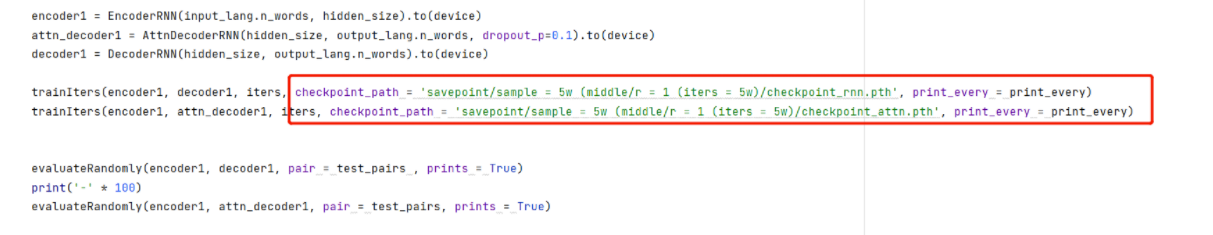
例如想导入reduce\_en-zh.txt，代码该为：open('data2/main/reduce\_%s-%s.txt' % (lang1, lang2), encoding='utf-8').\

下面介绍数据保存：

保存数据在savepoint文件夹中，第一级子文件夹指明使用哪个数据集，第二级子文件夹指明训练轮次和应当设置的iters值。



项目会自动保存savepoint，保存地址在此处设置：



需要注意的是，保存文件夹的要与导入数据集保持一致，且文件夹名称中的iters与代码设置的iters需要保持一致，否则会报错。

iters在此处设置：



例如，如果想进行如下的任务训练：

应当导入reduce\_en-zh.txt，且iters设置为20000