说明文档

代码分为三个部分：download\_data.py（下载数据），time\_preprocessing.py（处理答题时间过长或过短的文件），filtrating.py（筛选不合适的文件）。

1. download\_data.py  
    首先用一个爬虫程序，将index.txt和数据文件爬取到当前文件夹，数据文件按照index.txt中的路径名创建文件夹并存放，方便后续以index.txt为目录遍历数据文件。
2. time\_preprocessing.py  
    删除空文件。再以5Hz的频率计算出每个文件的时间，查看时间的分布情况，删除时间过长或过短的文件。
3. filtrating.py  
    以每个文件夹为单位，对文件进行预处理。
4. 检查文件中是否有缺失值。若有，则删除这一行数据。
5. 处理由于人为原因导致的数据异常。例如将手机放在桌子上答题，在图像上x,y,z轴数据波动不明显，则删除该类数据；还有图像波动特别大的，可能是答题者不是一直坐着或站着，身体带动手机有很大起伏，这种数据也要删除。