Homework10数据清洗说明文档

我的总体思路分为以下4部分：

1. 我写了一个爬虫程序将所有数据爬取到本地
2. 将前四步得到的文件列表进行分类，按文件名可分为6类（一共三个大类，每个大类分为健康和焦虑）
3. 之后考虑每个文件是否有缺失值，如有缺失值，则去掉该组数据。
4. 之后考虑将采集时间过长或者过短的数据过滤掉，由于采集频率是5HZ，所以只需要用数据条数除以5即可得到时间，然后我的标准是保留10分钟到30分钟的数据。