**数据预处理说明文档**

1. 写一个爬虫，首先爬取首页中的数据，对数据进行整理分类，分为六个不同类，六各不同的类放在六个文件夹，再对数据处理得到分为六类后属于某类的文件名，放在六个列表中，在根据列表中的文件夹名对数据进行爬取，放在文件夹中。
2. 对不同文件夹的文件进行统一处理
3. 每个文件中的数据集的长度不一，由采样频率5HZ,保留采取样本在10分钟和60分钟的文件。
4. 用数据集每个维度的平均值替换相应维度的空值。
5. 计算各个维度的方差，考虑可能设备故障和使用方法不当，删除方差小的文件。
6. 最后将保留的数据集再写入文件。