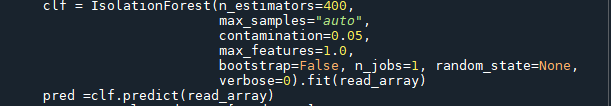
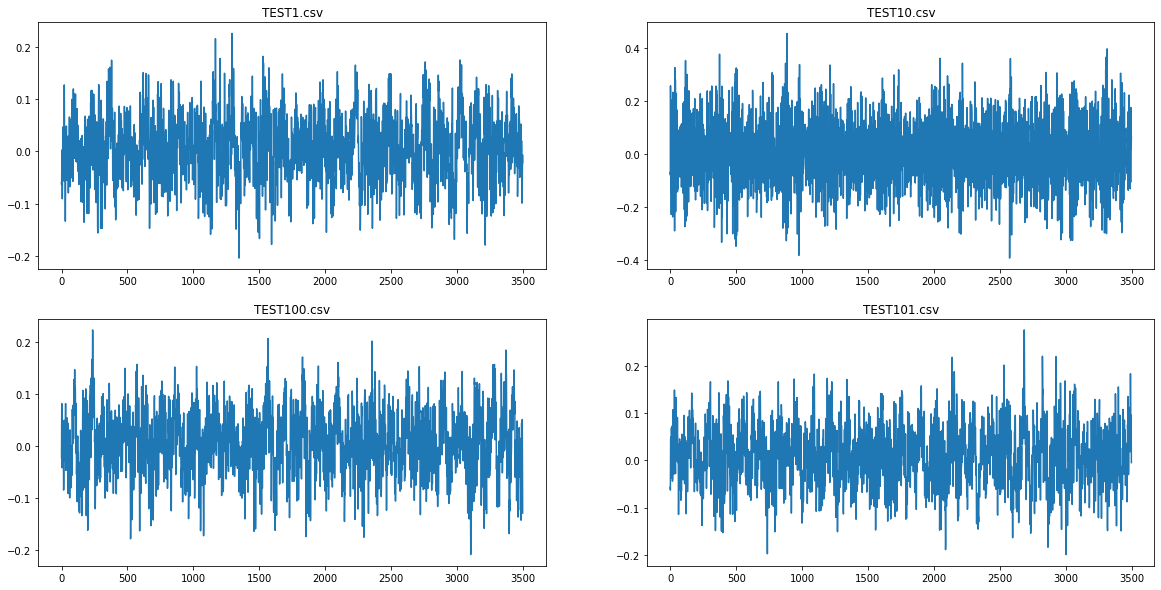
数据清洗：

使用独立森林算法，进行异常点检测

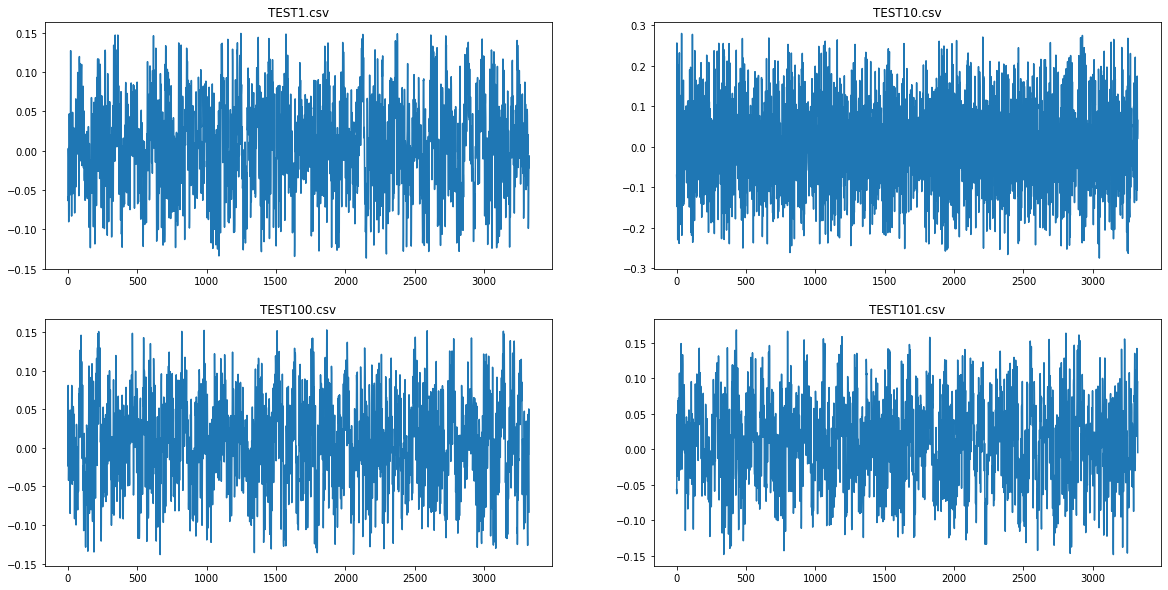
异常点的比例我设为了0.05



清洗之前



清洗之后

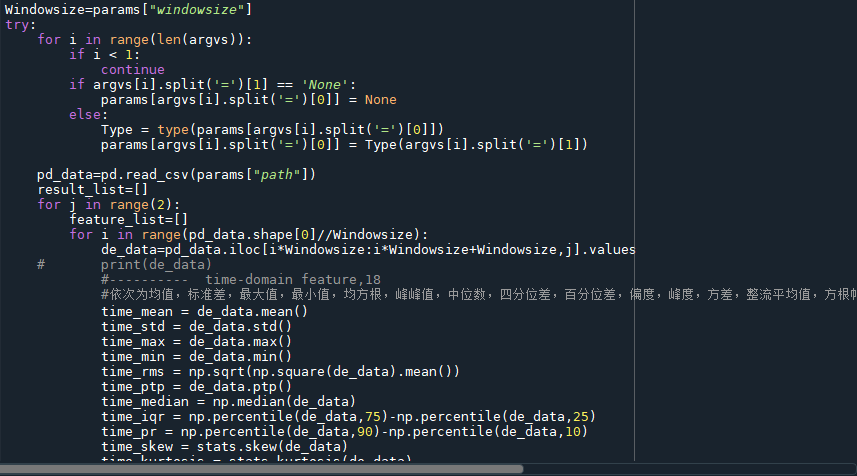


可以看粗话，一些明显突出的点被消除

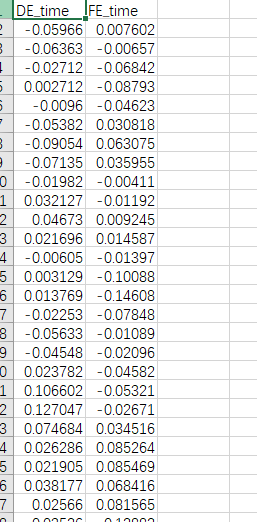
特征提取

时频域特征提取

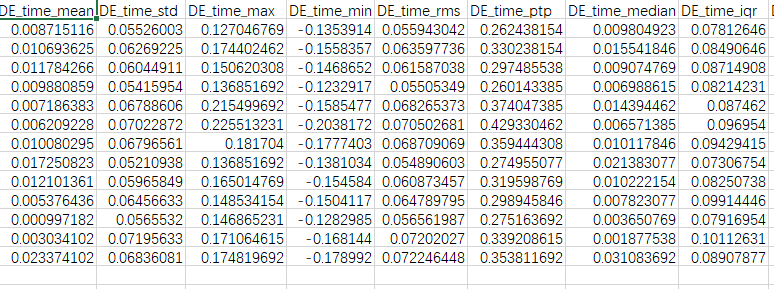
时间窗大小设为250



提取特征之前

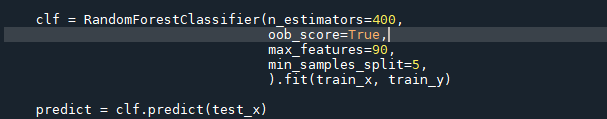


提取特征之后



训练

使用随机森林分类器进行分类



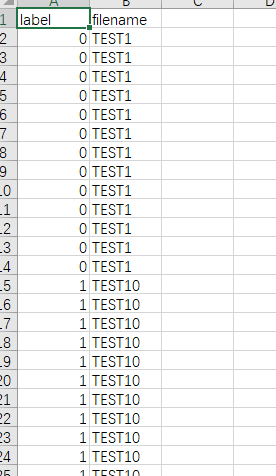
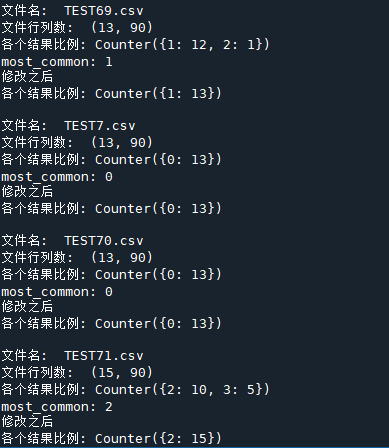
训练的结果尚可。

随后调用训练好的模型对测试集进行判断，得到结果。

通过模型，得到结果

示例

各个结果比例表示初步预测之后的结果，most\_common表示这个测试集概率最大的故障，然后将最大概率的故障当作本测试集的故障



模型优化：

1. 增加随机森林的树的数量，增加的400棵左右结果比较好。再多的话训练速度太慢，而且提升不大。
2. 数据清洗阶段调整异常点数的比例，从最初的0.05改到0.01，异常点比例设置太高，丢失的信息比较多。
3. 用一个测试集中概率最大的故障来代表这个测试集的故障。
4. 随机森林 最大特征数改为70，不设为90，可以略微提高点准确度