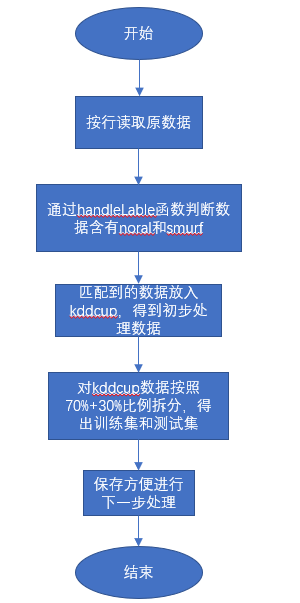
数据处理流程图：

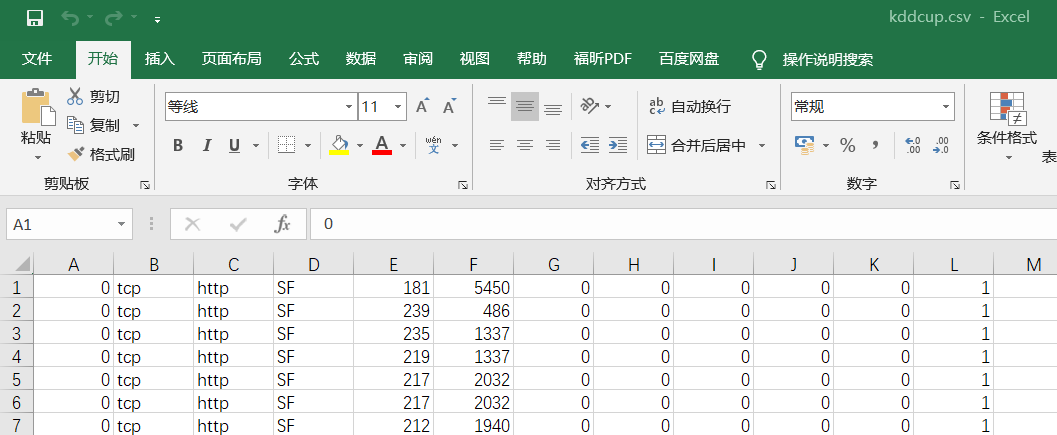


数据处理：

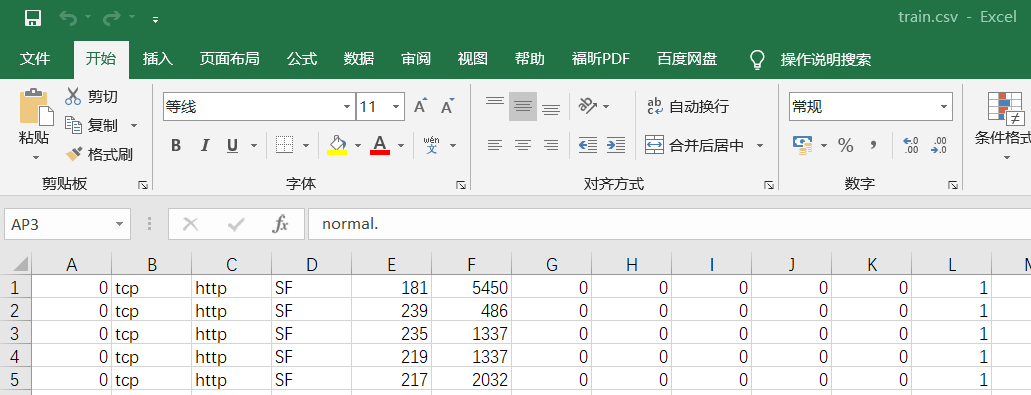
由于KDD1999数据集中包含数据种类较多，而本次实验只针对标记为normal和smurf的数据进行处理匹配，因此，对原文件为kddcup.data\_10\_percent的数据集首先进行读取匹配，匹配标记为normal和smurf的数据，并将其保存在kddcup.csv文件内以进行方便进一步处理和查看验证，再对kddcup.csv文件内的数据按照70%训练集和30%测试集的比例进行拆分，得到train.csv和attack.csv，方便程序进行进一步KD树构建和相关异常检测。

数据集结果：

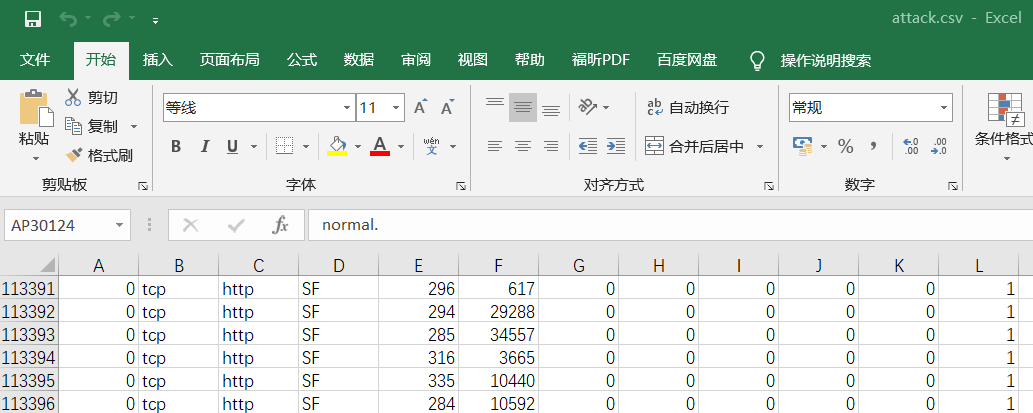
Kddcup初步处理数据集：



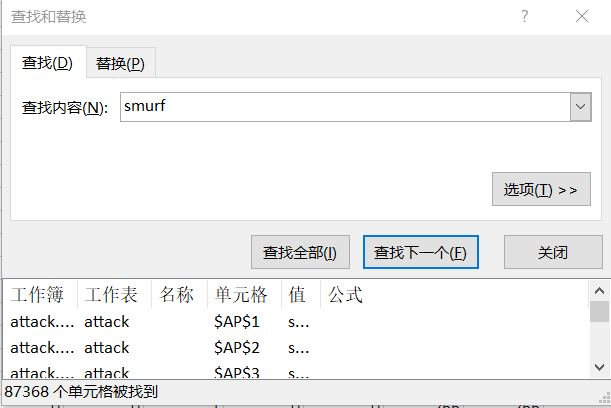
训练集train.csv:



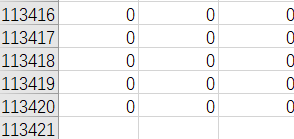
测试集attack.csv:



其中测试集中正常数据个数：



总个数



经计算训练集理论检测到的攻击比例为0.7703050608358314