Hit\_stop\_word.txt为停用词列表

特征.txt为爬取10页新闻得到的特征向量。

(1)爬取同花顺新闻财经数据

<http://fund.10jqka.com.cn/smxw_list/index.shtml>

此步骤主要使用的lxml,requests,pandas。其中request用于请求页面,lxml解析web页面，pandas用于存输数据。

（2）使用SVM同花顺财经新闻进行分类

新闻类别如下：利好，利空，中性

此步骤主要使用jieba, scikit-learn,其中jieba可以用于中文的分词（当然你可以使用其他中文分词包），可以使用scikit-learn中的svm对新闻数据进行分类。