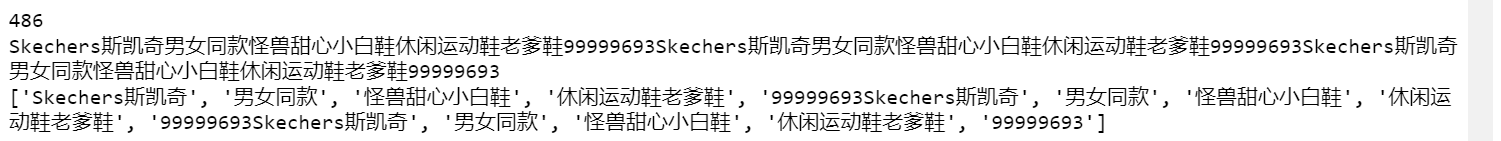
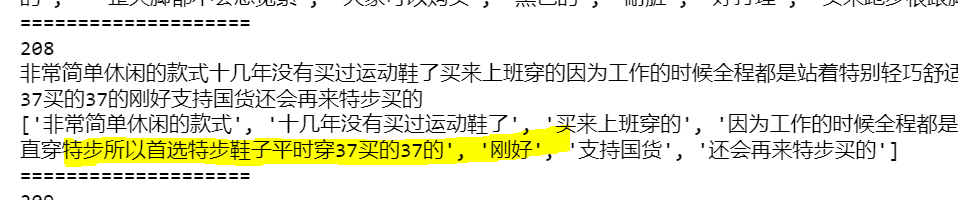


加油 中国分开了，需要在最后加分隔符|的位置写一个类似jieba add\_word免除分词的功能

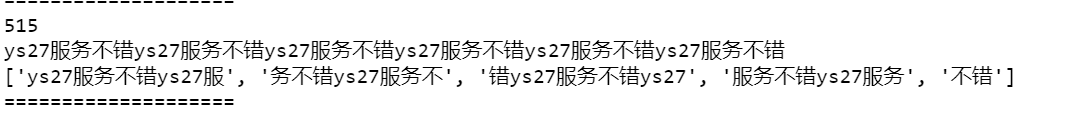
纯数字和纯英文的短语不要了

一个中文字的如果不是key\_words里的字也可以不要





一条评论里重复的短语也可以删了



建立垃圾句子库，在垃圾句子库里的句子也不要了（或者更高级点的垃圾句子分类模型）

奶奶->奶的去叠字问题