测试样本--

文本预处理模块：

1. 去除非中文部分
2. 分词

词典模块构造Sentiment：

1. 情感词库构造

情感词库实例：



情感分类有7大类，21类：

***乐***：快乐(PA),安心(PE)

***好***：尊敬(PD),赞扬(PH),相信(PG),喜爱(PB),祝愿(PK)

***怒***：愤怒(NA)

***哀***：悲伤(NB),失望(NJ),内疚(NH),思念(PF)

***惧***：慌张(NI),恐惧(NC),羞愧(NG)

***恶***：烦闷(NE),憎恶(ND),贬责(NN), 妒忌(NK),怀疑(NL)

***惊***：惊奇(PC)

情感极性：0表示中性，1表褒义，2表贬义

情感极性和情感分类的关系：

褒义

乐

好

贬义

怒

哀

惧

恶

中性

惊

a.选择词语，情感分类，强度，极性(积极，中性，消极)

b.将情感分类，极性转化为分类特征：

乐，好，怒，哀，惧，恶，惊

(s1，s2，s3，s4，s5，s6，s7)

情感分类

褒义，贬义，中性

（p1， p2， p3）

极性分类

情感分类向量

极性分类向量

c.将词语作为key，将情感分类，强度作为value，构造情感词典

{词语：[情感极性\*情感强度,情感分类\*情感强度]}

例如：/Users/zhaobowen/Desktop/Screen Shot 2018-07-02 at 3.53.53 PM.png

转化为：{‘脏乱’:[ [0,0,7], [0,0,0,0,0,7,0]]}

2）否定词词库negative

3）程度副词词库Degree：

为每个程度副词强度打分

构造程度分词词典：{程度副词：程度值}

情感向量计算模块

1.情感强度计算模块：

初始情感分类向量sent：（0，0，0，0，0，0，0）

初始化情感极性向量polar：（0，0，0）

初始权重：W = 1

while 循环遍历句子中的每一个词word：

条件判断：

1. 如果word 在否定词库中，W\*(-1)
2. 如果word在程度副词库的key中，W\*Degree[word]
3. 如果word在情感词典的key中

更新情感极性向量ploar += W\*Sentiment[word][0]

更新情感词向量 sent += W\*Sentiment[word][1]

权重W重置为1

2.句子情感分类：

1. 句子极性分类：选择情感极性向量中polar最大值者
2. 句子情感分类：

if 句子极性为积极：从[乐, 好]中选择值大的类别作为句子的情感分类

if 句子极性为中性：选择 ‘惊’作为情感分类

if 句子极性为消极：从 [怒，哀，惧]中选择值大的作为句子的情感分类

存在的一些问题：

1. 句子中没有出现情感词，需要添加默认情感分类
2. 否定词、或肯定词修饰的词语不在情感词典中:

比如：我不想买这辆车，因为它开起来不是很舒适。

按照规则，舒适前面被两侧否定词修饰，则这句话变成了积极的情感

1. 句子中虽然没有情感词，但有否定词：这辆车性能不快—消极情感
2. 句子中如果仅出现一类情感词，比如属于乐、好的情感词，由于否定词的存在，使得句子的极性为否定，但由于消极情感词汇(属于怒，哀，惧的情感词)并未在句子中出现，因而无法对句子进行情感分类

上述问题主要来源于两点：

1. 情感词汇较少，且情感向量是稀疏的，导致一个句子的情感分类向量也是稀疏的。
2. 规则较为简单，且没有对句子进行句法分析（否定词、程度副词是否修饰的是情感词）