

基于大连理工情感词典的情感分析和情绪计算

情绪包括anger、disgust、fear、sadness、surprise、good、happy

<https://www.flyai.com/article/679>

情绪	对应情感
Happy	PA,PE
Good	PD,PH,PG,PB,PK
Surprise	PC
Sad	NB,NJ,NH,PF
fear	NI,NC,NG
Disgust	NE,ND,NN,NK,NL
Anger	NAU (原NA)

以影评为例：

"原著的确更吸引编剧读下去，所以跟《诛仙》系列明显感觉到编剧只看过故事大纲比，这个剧的编剧完整阅读过小说。配乐活泼俏皮，除了强硬穿越的台词轻微尴尬，最应该尴尬的感情戏反而入戏，故意模糊了陈萍萍的太监身份、太子跟长公主的暧昧关系，整体观影感受极好，很期待第二季拍大东山之役。玩弄人心的阴谋阳谋都不狗血，架空的设定能摆脱历史背景，服装道具能有更自由的发挥空间，特别喜欢庆帝的闺房。以后还是少看国产剧，太长了，还是精短美剧更适合休闲，追这个太累。王启年真是太可爱了。"

- 首先利用jieba库分词（去除停词，留下情绪词、否定词、副词）
 1. 情绪词存在于情感词典当中
 2. 否定词存在于否定词表中
 3. 副词为程度副词，包括非常、丝毫等，对应不同的权重值
- 接下来对各情绪特征统计分数
 1. 首先对每个情绪词（非副词、非否定词）计算情绪值 V_i
 - 例如，对于上文“活泼”，情感词典中为分类为“PH”，对应情绪特征“GOOD”，对该词polarity进行讨论
 1. 若 $Polarity_i = 0$ ，则 $V_i = 0$
 2. 若 $Polarity_i \neq 0$ ，则 $V_i = intensity_i$
 2. 对情绪词 i 和情绪词 $i - 1$ 之间的否定词和副词，定义权重 $W_i = 1$ ，按以下情况讨论
 1. $W_0 = 1$
 2. 若没有否定词或副词， $W_i = 1$
 3. 若只有否定词，定义否定词个数为 t ，则 $W_i = (-1)^t$

4. 若只有副词，则 $W_i = \prod_k C_k$ ，其中 C_k 表示第 i 个程度副词的权重值（从程度副词表中得到）
5. 若既有否定词，也有程度副词，需要看先后关系
 - 若否定词在程度副词之前，例如“不是很”
则 $W_{i+} = \frac{C}{2}$, C 为副词的权重值
 - 若否定词在程度副词之后，例如“很不是”
则 $W_{i+} = -C$
3. 对于该单词的情绪特征（活泼对应Good）， $V_{sentiment_i+} = W_i * V_i$
4. 对于所有情感，计算其 $V_{sentiment}$ ，最高者即为该文本对应情感。