

文献5复现

1. 对于数据集中每个句子，进行一次分词，得到数据格式 情绪类别 句子的词序列 句子
2. 对于每一种情绪，把训练集中的一条数据进行处理，计算一个句子的每个词的先验

$$P(w_i|c_j) = \frac{w_i \text{ 在该情绪下的所有推文中出现的次数} + 1}{\sum_{q=1}^n w_q \text{ 在该情绪下的所有推文中出现的次数} + n}$$

问题：这里如何用每个句子去更新 $P(w_i|c_j)$ ，和步骤4更新的公式一样

3. 对于一个新出现的句子 t ，对每一个情绪类别 计算 $P(c_j) \prod_{q=1}^n P(w_q|c_j)$ ，最大那个就是该句子的分类
4. 增长的方法，即在训练集中部分作为更新 $P(w_i|c_j)$ ，但是更新的公式是？