贝叶斯文本分类器与概率论

**实验目的**

* 通过贝叶斯算法对评论按照正负情感进行分类

**模型设计**

* 求评论D属于类别C的概率，即
* 因即求 
* 即
* 其中k为类C下所有的词语数量，C为情感类别（正面或者负面），D为评论；
* 可先求得
* 因先验概率，可计算训练数据中类别C出现的概率，C条件下Tk的贡献度
* 根据后验概率，计算出评论中的贡献度之和(可省去分母全概率公式的计算，判断正面或者负面时，分母一致，只判断分子大小即可)，根据贡献值大小，判断属于正面或者负面

**实验结果**

* IMG_256
* 负面召回率低，进行代码检查，尝试进行去除样本中非中文字符；检查训练数据

**训练样本数**

* 正面500条，负面500条

**采用模型**

* 多项式模式

**样本空间**

* n{0<=n<1}，任意特征词语对于评论情感正负的贡献值

**基本事件**

* 各个特征词语对于情感值的贡献度，如
* 词语“不错”对于正面情感的贡献度:0.0107402511566
* 词语“不错”对于负面情感的贡献度: 0.00348374253484

**独立事件**

* 词语A对于正负面评论的贡献值，与词语B对于正负面评论的贡献值相互独立