**邮件分类的思路分析**

特征：邮件的内容

标签：正常邮件或垃圾邮件

训练集：4800封正常邮件和3200封垃圾邮件

测试集：1200封正常邮件和800封垃圾邮件

# 提取特征

使用jieba库对邮件内容进行分词处理，然后将文本转换为数值特征。

# 训练模型

根据训练集中的邮件的内容和对应的标签（正常邮件或垃圾邮件）来训练一个分类模型。

p(正常邮件)=0.6：邮件是正常邮件的概率是0.6。

p(垃圾邮件)=0.4：邮件是垃圾邮件的概率是0.4。

# 模型评估

在模型训练完成后，使用测试集来评估模型的性能。包括准确率、精确率、召回率和F1分数。

# 概率计算

p(正常邮件∣邮件内容)= p(邮件内容∣正常邮件)/p(邮件内容)

p(垃圾邮件∣邮件内容)= p(邮件内容∣垃圾邮件)/p(邮件内容)

由此推出：

p(正常邮件|邮件内容)=0.3：邮件是正常邮件的概率是0.3。

p(垃圾邮件|邮件内容)=0.2：邮件是垃圾邮件的概率是0.2。