结论分析：

本次实验所有模型采用的训练集和测试集相同，其中训练集容量为90000条弹幕，测试集容量为10000条弹幕。采用的模型有朴素贝叶斯模型、支持向量机模型和梯度提升树模型。其中朴素贝叶斯模型包括高斯模型、多项式模型和伯努利模型，支持向量机模型内核包括线性核和径向基核。

整体来看，支持向量基和梯度提升树模型的效果要优于朴素贝叶斯模型。几种朴素贝叶斯模型中，伯努利模型整体效果优于其他两种。当特征为500时，高斯和伯努利模型中one-hot独热码、wf词数、tf词频和tf-idf词频-逆文本频率编码正确率最高，随着特征数量的提升，在4000特征时，正确率最低。高斯模型中wf词数编码效果整体最好，其次tf-idf编码。One-hot和wf编码在特征数量少的时候效果优于tf词频、tf-idf调频-逆文本以及word2vec词向量编码。word2vec词向量编码随着特征数量的变化，正确率变化并无规律，数值差别不大，且整体正确率低于其他4种编码。除高斯和伯努利模型外，多项式模型在500特征时各编码正确率最低，在4000特征时正确率最高。其中效果最好的是tf词频编码，其次是tf-idf词频-逆文本频率编码，one-hot独热编码效果最差。随着特征数量的提升，多项式模型的正确率提升较快，在3000特征时优于高斯模型，在4000特征时优于贝叶斯模型。

支持向量机模型和梯度提升树模型整体比朴素贝叶斯模型效果好。其中使用线性核函数模型效果优于径向基核函数模型。线性核函数支持向量机one-hot独热码正确率最高，且在4000特征时，弹幕识别正确率达到93.3%。正确率真相仿的是wf词数编码。Tf词频和tf-idf词频-逆文本频率编码效果一般，且都优于朴素贝叶斯模型。Word2vec模型效果最差，正确率接近58%且分布平均。径向基核函数模型中效果最好的是tf模型，随着特征数量的增加，效果变差。其中word2vec词向量编码效果明显优于线性核函数模型中的word2vec词向量编码效果。梯度提升树模型在特征数量变化时，变化效果不显著。效果最好的是word2vec词向量编码，效果最差的是one-hot独热码。