互联网虚假新闻检测

# **摘要**

随着互联网的飞速发展，家家户户都逐渐牵上了网线，凭借着网络的便利性，外界信息得以轻松的得知，对于百姓们熟知的新闻信息的获取也能在须臾之间完成。然而任何事物的发展都具有两面性，

在这个社交媒体和互联网如此发达的世界中，我们接触到的各种新闻和文章，但其中一些新闻和文章是假的，是为了好玩或出于恶意而制作的，这对社会是非常有害的。想对媒体产生的新闻进行鉴别，实现对虚假新闻和真实新闻的检测分类识别，以便用户对获取到的新闻进行清晰的认知，排除虚假新闻对用户的影响。

# **简介**

在如今网络高速发达的时代，互联网虽说使得信息的产生和获取更加便捷，也滋生了不少虚假新闻扰乱民众的视听，倘若放任不管甚至可能导致谣言泛滥成灾，最终影响到现实社会秩序。在某种程度上假新闻的是一个微观领域问题，它和谣言分类，事实判

断，标题党检测，垃圾内容挖掘等都比较相似

在宏观上说都属于内容质量的领域，所以很多方法其实是通用的框架。

本次课程设计通过三种方案解决假新闻检测的问题。

1. TF-IDF+被动攻击算法
2. NaiveBayes
3. MLP模型

# **TF-IDF+被动攻击算法**

1. TF-IDF Vectorizer。

TF（词频）：一个词在文档中出现的次数就是它的词频。较高的值意味着某个词比其他词出现的频率更高，因此，当该词是搜索词的一部分时，该文档是一个很好的匹配项。

IDF（逆文档频率）：在一个文档中多次出现但在许多其他文档中也出现多次的词可能无关紧要。IDF 衡量一个词在整个语料库中的重要性。

TF-IDF Vectorizer将原始文档集合转换为TF-IDF特征矩阵。

1. 被动攻击算法（Passive Aggressive Classifier）

被动攻击算法是在线学习算法，用于从大量的数据中进行学习。在样本分类正确且模型对可能性的预测准确时，模型不做调整（这里体现出了被动）。在样本分类正确但模型对可能性的预测有失偏颇（不太准确）时，模型做出轻微的调整。在样本分类错误时，模型做出较大的调整（体现出较强的“攻击性”）。是从包含政治新闻的数据集中提取数据，用TF-IDF Vectorizer将其转换为TF-IDF 特征矩阵，运行被动攻击算法，拟合模型。

1. 混淆矩阵

在机器学习中，如果输出应该生成两个或者两个以上的类，则用于性能度量，有四种可能结果：

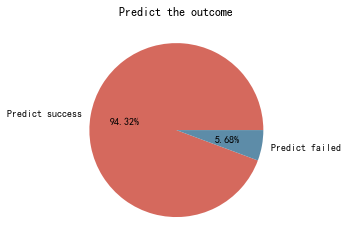
真阳性-本身是对的，你当成是对的了

假阳性-本身是错的，你当成是对的了

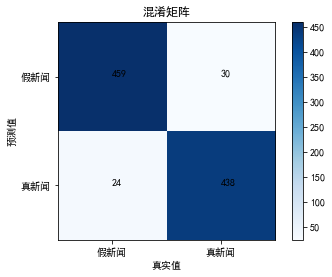
真阴性-本身是错的，你当成是错的了

假阴性-本身是对的，你当成是错的了

1. 算法结果



预测结果



混淆矩阵

# **NaiveBayes**

朴素贝叶斯算法的特征提取使用了sklearn的特征提取库feature\_extraction.text。

整体流程：预处理后生成文本分词表，再通过朴素贝叶斯专门的预处理流程进行预处理，再代入算法计算。首先通过数据清洗\预处理函数对初始数据进行清洗，数据清洗函数在dataload.py文件的preprocess类中，先调用preprocesingdata\_news函数对数据统一清洗（提词分词）,再分别调用processingdata\_bayes和processingbert函数将初步清洗的数据分别转换成训练相应的模型需要的格式。朴素贝叶斯算法的格式主要为列表和字符串，每一个词语之间用空格分开，每一个句子都是单独的列表。标签同样也是列表存储。

朴素贝叶斯算法流程图如下：



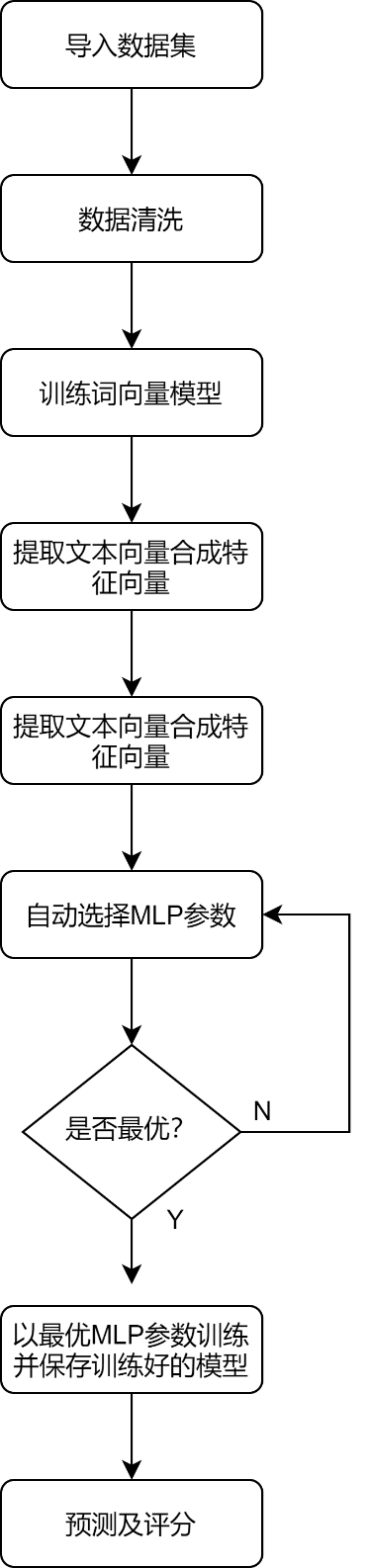
算法结果：

# **MLP模型**

1. 预处理proccess；这个阶段要进行数据导入，数据清理和分词工作，此外还要生成词向量模型，再根据得到的词向量模型构建样本的特征向量。
2. 模型参数自动调优 optimize；
3. 模型训练 fit；本项目将采用多层感知机的机器学习模型，该模型是一个多分类的线性分类模型。
4. 预测 predict；
5. 分析优化 analyze；

由于中文分词、训练词向量、模型参数选择、模型训练这四个部分耗时较大，为了避免不必要的时间浪费，我将分词结果、词向量模型、模型参数、训练好的MLP模型分别保存，方便下一阶段直接使用，以实现模型的持久化。

总体的算法流程图描述如下：



算法结果：

.