**假新闻分类**

一、问题描述

文本分类是自然语言处理 (Natural Language Processing, NLP) 应用领域中最常见也最重要的任务类型，在搜索引擎、问答系统、会话系统等等重要的信息处理系统中应用非常广泛。循环神经网络（Recurrent Neural Network, RNN）是一类以序列数据为输入，在序列的演进方向进行递归且所有节点（循环单元）按链式连接的递归神经网络。本项目将利用循环神经网络实现假新闻分类。

1. 参考阅读
2. Recurrent Convolutional Neural Networks for Text Classification

2. LSTM：《Long Short-Term Memory》

3. GRU：《Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation》

三、数据集合

假新闻数据集有FakeNewsNet、BuzzFeedNews、LIAR等。本次我们使用Kaggle的Fake and real news dataset数据集，它有两个文件，一个是真实新闻，一个是假新闻(都是英文)，总共有23481条“假”推文和21417条“真实”文章。

**数据集下载方式：**

在以下链接下载：https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset

四、建议方法

**基本方法**：搭建RNN完成假新闻分类。

进阶方法：尝试搭建LSTM、GRU其中一种模型完成假新闻分类，并且对结果进行准确性评估。

五、评估方法

真假新闻分类的准确率。