文件说明：

首先下载词向量。代码中使用的词向量是人民日报的 300d word+Ngram 词向量。

Loaddata 内部封装了 word_emb()函数，调用以加载词向量。Preprocess 的作用是为了处理文章，得到每个汉字的索引，训练集和测试集分别存放在 data.json 和 valid.json 里面，包含 textdata—文章内容编码、labeldata—label 数量信息以及 ce—用于计算 crossentropy 的向量。

Model.py 是 CNN 模型，train.py 可用于重新训练模型或加载已经训练好的模型。提交时的模式为重新训练模式，如要加载模型可将##FOR TEST 以下的部分的注释删掉。CNN 的保存模型是每一个 epoch 保存一次，加载的时候修改下方 epoch 的值可选择究竟加载哪一个模型。附件中提供的是第 8 个 epoch 的模型，它是报告中提到效果最好的 CNN 模型。

Kera.py 是 RNN 模型，mlp.py 是 MLP 模型，都可用于重新训练或加载已训练好的模型。如要加载模型，则需注释掉 model.fit，并把标注 FOR TEST 的部分的注释去掉（包括 model 中的 load_weight）。设置的保存模式为验证集 loss 最小，每当验证集 loss 变小就保存一次，否则不保存。RNN 提供的可加载模型是正确率最高的版本。MLP 没有提供可加载模型，因为它太大了……

注：文件比较大的原因是我的 input_length 取了 1500…这样确实比别人的的文件大很多……取这个值也是我经验不足导致的，我以为 input_length 要覆盖掉绝大部分文章，但其实不用。在这里对可能给助教带来的麻烦表示深深的歉意！