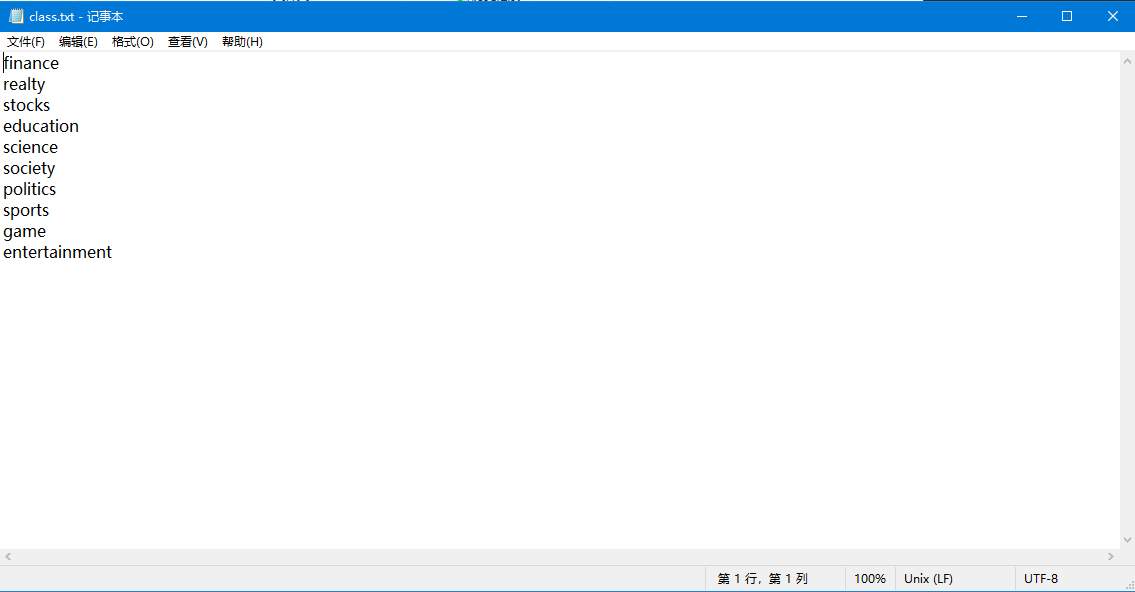
摘要分类模型使用说明

模型的使用分为三个步骤:

**1. 数据集的制作**

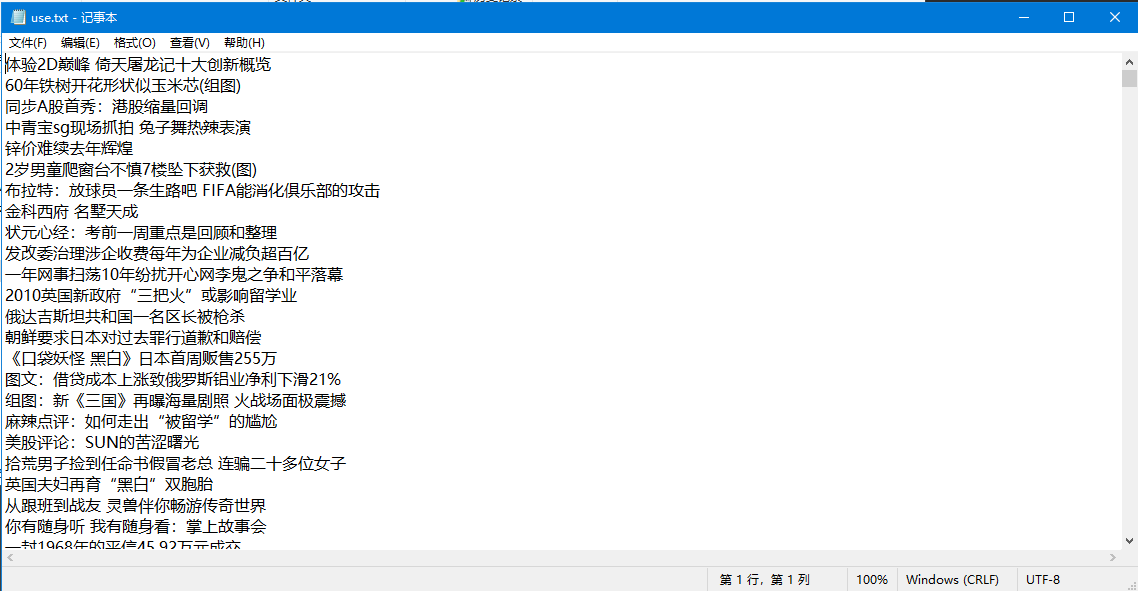
**>>>**首先确认数据摘要分为几类, 将类名写到文件”THUCNews/data/class.txt”中.

例:



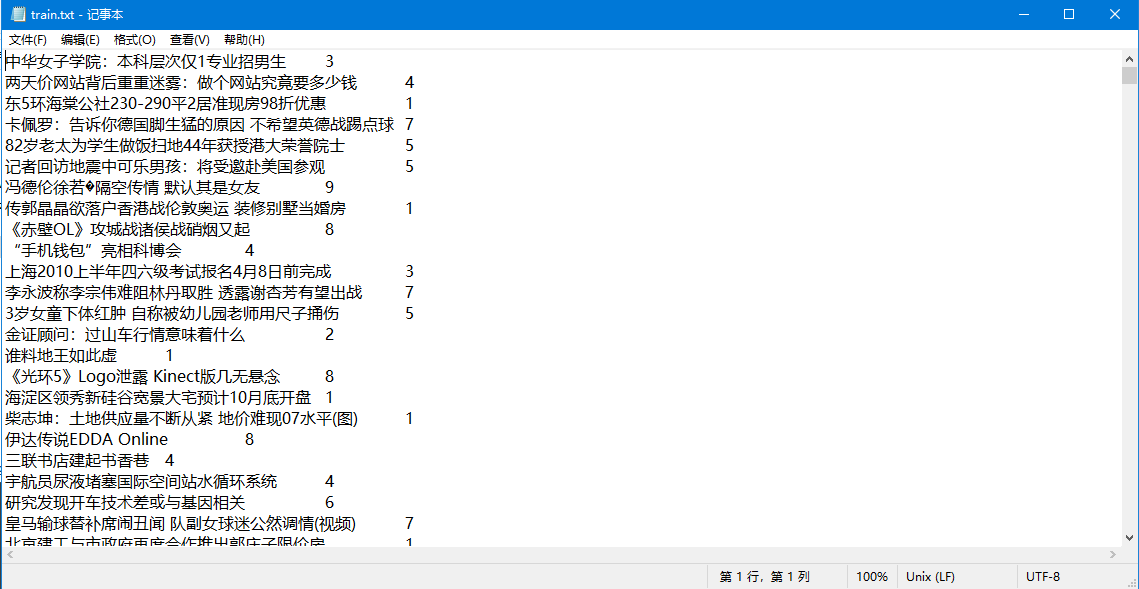
**>>>**将摘要以每行一条的形式制作成txt文件, 并划分成”训练集”(命名为”train.txt”), “测试集”(命名为”text.txt”), “验证集”(命名为”dev.txt”)和”使用集”(命名为”use.txt”). 文件均放到”THUCNews/data”文件夹下, 请严格按照规则命名!!!!

例:



**>>>**给训练集, 测试集和验证集打上标签, 按照前面的分类在每一个摘要后面隔一个tab键写下对应的数字. 例如一条新闻属于教育类, 在前面的文件中教育类在第四个, 所以在该条新闻题目后标记4

例:



data文件夹下目前存放的是需要填入文件的空白同名文件, 无法运行模型. 制作好数据集后替换就好.

如果对数据集格式还是不清楚, 可以查看文件夹data\_example中的文件.

**请注意:**

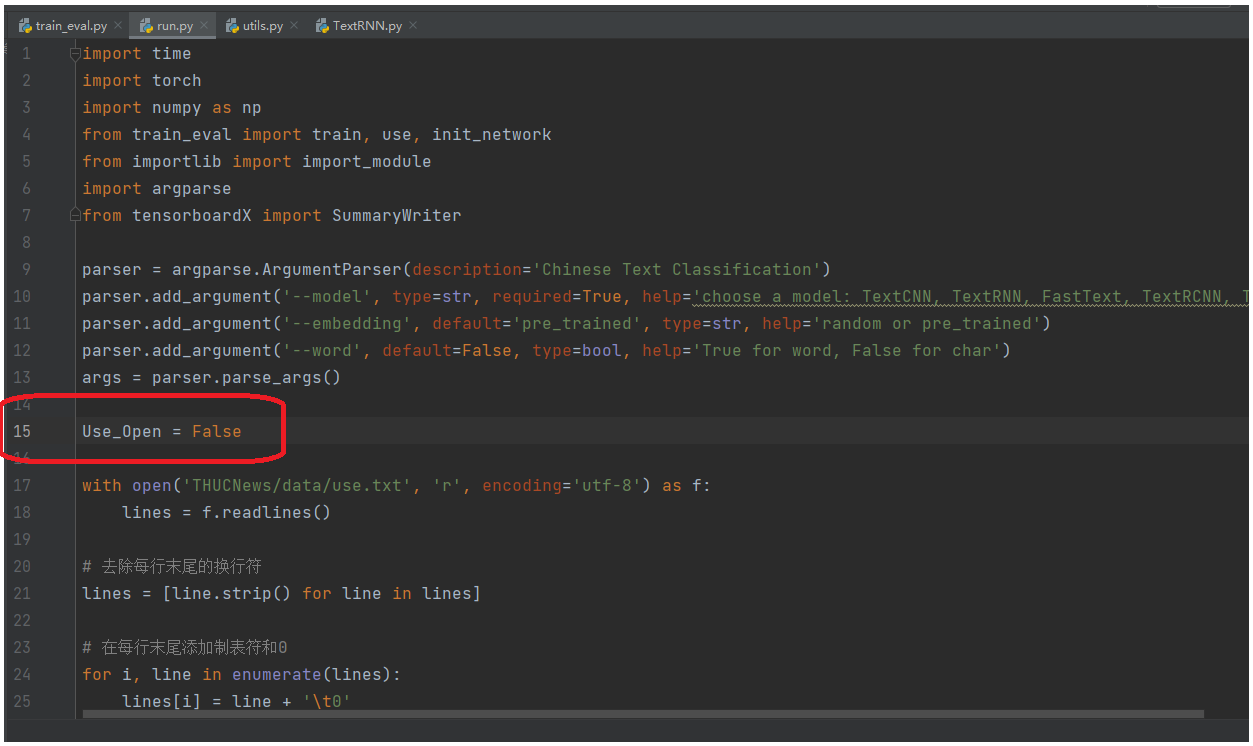
**1. 请确保摘要的类型在训练集, 测试集和验证集里都有包括, 例如分类为7类, 则三个数据集中有且仅有7类的摘要.**

**2. 分类名不要写中文.**

**2. 训练模型**

**>>>**“run.py”为模型入口, 运行该文件启动模型.

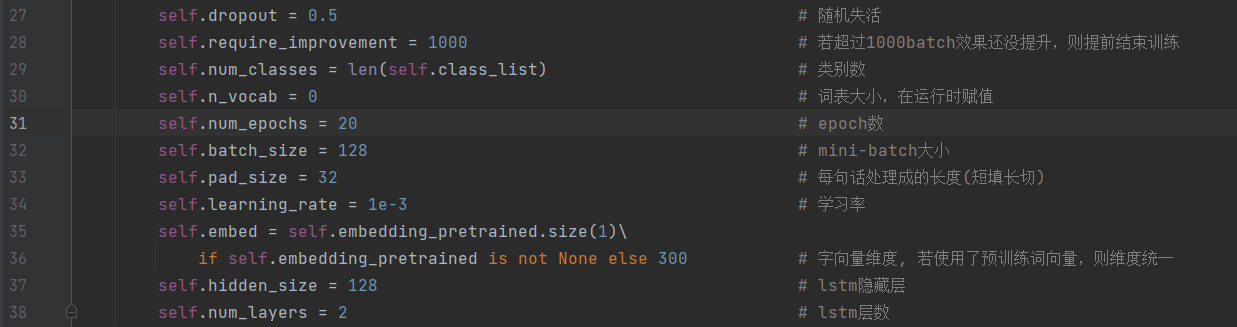
run.py文件中的第15行的Use\_Open, 赋值为True表明开始使用模型, 为False表明训练和测试模型, 根据需要对参数进行调整.



**>>>**模型参数设置在TextRNN.py文件中, 可以根据模型的训练效果进行调整.

pad\_size表示读入模型的字的数量, 根据摘要的长短进行重写.(例如摘要平均长度为10个字, 则pad\_size就设置为10)

(batch\_size改了不知道会不会报错, 如果报错再改回来就好)



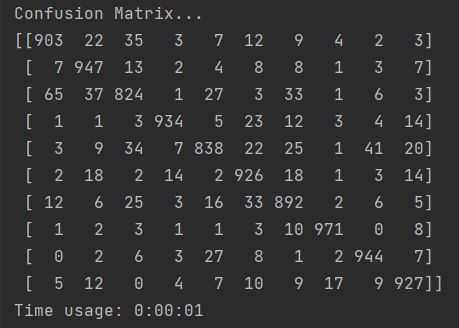
**训练结果说明:**

模型训练完成之后, 会生成一个分类表和一个混淆矩阵:

**>>>**分类表:



**>>>**混淆矩阵:

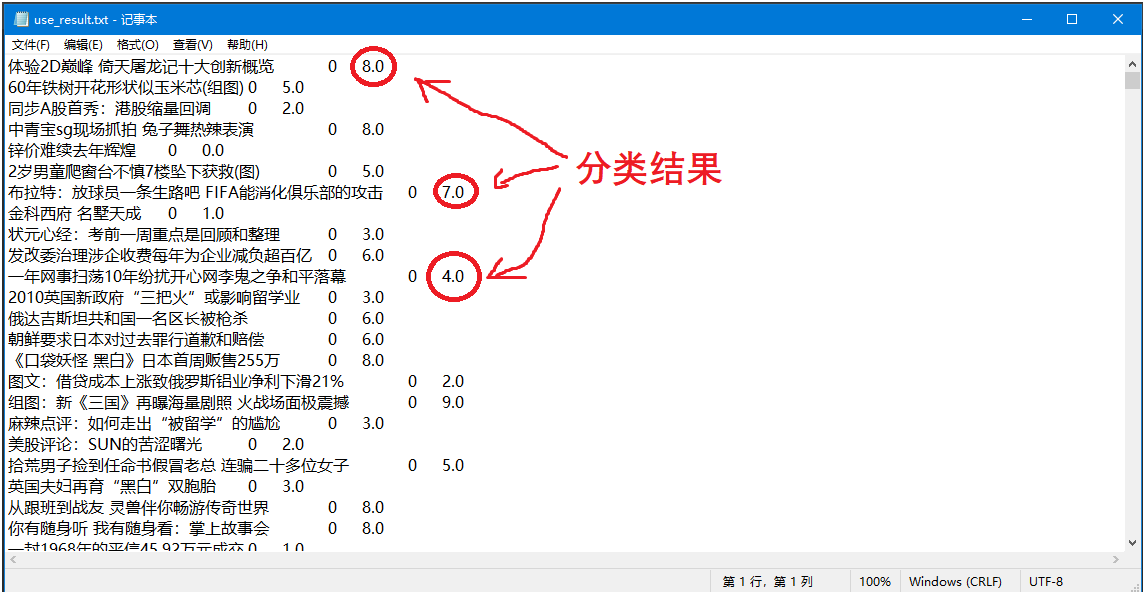


表示每个类别模型的识别情况, 例如第一行表示第一个类别的1000条数据, 有903个识别成第一类, 22个识别成第二类, 35个识别成第三类………

**3. 模型使用**

首先将Use\_Open设置为True.

运行模型后会在data文件夹下生成”use\_result.txt”文件, 该文件会将use.txt文件中的摘要自动分类, 如图:



**注意: 请先训练模型再使用, 不然会报错!!**

模型使用的大致说明到此结束, 如果还有不清楚的地方请和我联系.

2023.09.13