1. **读取数据**

本次竞赛任务为细粒度角色情感预测任务，训练数据中包含id，content，character，emotions等数据列，其中id通过script\_id，scene\_num,以及sentence\_num组成，emotion是情感识别结果，按顺序分别对应爱、乐、惊、怒、恐、哀的情感值，读取数据代码示例如下。为更好的联系上下文对目标文本进行分类，在切分训练数据与验证数据时请尽量以剧本和场次为单位进行切分。

file\_path = ('train.txt')

script\_ids = []

scene\_nums = []

sentence\_nums = []

ids = []

contents = []

characters = []

emotions = []

index = 0

with open(file\_path,'r',encoding='utf-8') as f:

for line in f.readlines():

if index > 0:

item = line.replace('\n','').split('\t')

id,content,character,emotion = item[0],item[1],item[2],item[3]

script\_id,scene\_num,sentence\_num = id.split('\_')[0],id.split('\_')[1],id.split('\_')[3]

script\_ids.append(script\_id)

scene\_nums.append(scene\_num)

sentence\_nums.append(sentence\_num)

ids.append(id)

contents.append(content)

characters.append(character)

emotions.append(emotion)

index += 1

验证数据的格式和训练数据基本一致，只是去掉了最后一列emotions。

1. **模型定义与训练**

为保证模型效果，请尽量选择BERT等预训练语言模型，并建议联系同一个剧本场次的上下文，通过多分类或回归方法，对目标情感进行预测。

当前的Baseline score是0.69535，截止到2021年9月23日下午19时，在A榜上可以排到第6名。

1. **输出结果**

输出结果的格式与训练数据的格式一致。