

一、实验目的

学习对话情感分析模型，能够较准确预测出对话包含的情感。

二、实验步骤与过程

2.1 情感分析案例

2.1.1 数据集介绍

(1) 训练样本 train_data.csv

该数据集分为了 ID, Text, Labels 三种数据，共有 8415 组对话，每组对话中的句子有各自的情感标签。本次实验随机取数据集的 70%作为训练集，10%作为验证集，20%作为测试集。

(2) 测试样本 test_data_new.csv

该数据集分为了 ID, Text, Labels 三种数据，共有 1000 组对话，每组对话中的句子除最后一部分外有各自的情感标签。在测试阶段将全部数据作为测试集。

2.1.2 数据处理

(1) 训练阶段

在该阶段试图将数据处理成为与 dailydialog 数据集中相同的格式，即每段对话为一组，每组中第一列为 0 与 1 交替代表对话者，第二列存入将原样本以 ' __eou__ ' 分割后的每一句话，第三列为该句话的情感，每组之间以一个空行相连接。同时由于原数据集中的 Labels 部分为数字，需要将其按照字典转换为对应的真实情感。

对于划分好的训练集、验证集、测试集分别采取以上操作，存入 /dataset/newdata 中。

(2) 测试阶段

由于测试样本中缺少每段对话最后一句话的情感标签，拟采用将原 Labels 增长 1 位，为该段话随机生成一种情感，其格式与训练阶段数据相同。

2.1.3 训练阶段展示

训练参数信息如下：

Epoch: 共 15 轮

硬件资源: 单 GPU

学习率: $1e-6$

训练时间: 15h

训练过程中保留 acc 最高的轮数的权重，存入到 /newdata_models/roberta_large/pretrained/no_freeze/emotion/1.0/model.bin 中。

```
Final Fscore ## test-accuracy: 0.6936297912972309, test-fscore: 0.6842562638652202, test_epoch: 13
```

在本次实验中，训练结果最好的一轮为第 13 轮，在训练阶段测试集上的 acc 达到了 0.69。

2.1.4 测试阶段展示

根据模型预测出的每句话的预测值，提取出每段对话中最后一句话的情感预测值，与真实结果做对比，准确率达到了 0.675。

```
, accuracy:0.675,
```

三、实验结论

3.1 实验结论

在数据集方面，本次实验采用的数据集按照 dialydialog 数据集格式转化，准确率相对于其他数据集较高，但由于在不同数据集上准确率分布差异较大，因此该模型泛化能力不强，需要加以改进。

当数据集规模较大时，模型所需要计算的资源会提高，训练效果会受到计算机配置影响，因此，在模型训练过程中只能降低参数要求，延长模型训练时间来完成模型的训练，同样导致准确率下降。参数设置不好容易造成过拟合的情况。