

# Paraphrase Identification Demo

本Demo基于BERT搭建起一个简单的匹配模型，输入两个文本，输出结果是否互为复述。

## 目录结构：

- /Model
  - MatchModel.py 模型文件，包含了输入输出的运算过程（其中包含了其他预训练模型的匹配模型，本demo只使用了BERT，其余可以忽略）
- /utils
  - logger.py 日志记录
  - classification\_metrics.py 评价指标
- all\_dataset.py 读入数据文件，将数据转换成pytorch的dataset格式
- parser1.py 参数设置文件，可以通过命令行传入
- Train\_baseline.py 训练代码，通过train\_lcqmc.sh调用
- train\_lcqmc.sh 启动训练脚本（入口，传入参数，调整参数）
- interface.py 训练完成后，使用训练好的模型进行预测的接口（传入模型路径即可）
- requirements.txt 运行环境所需包（pip3 install -r requirements.txt）
- readme.md 本文件，说明

## 使用方法：

- 训练模型：

```
sh train_lcqmc.sh
```

可以通过文件修改训练超参数

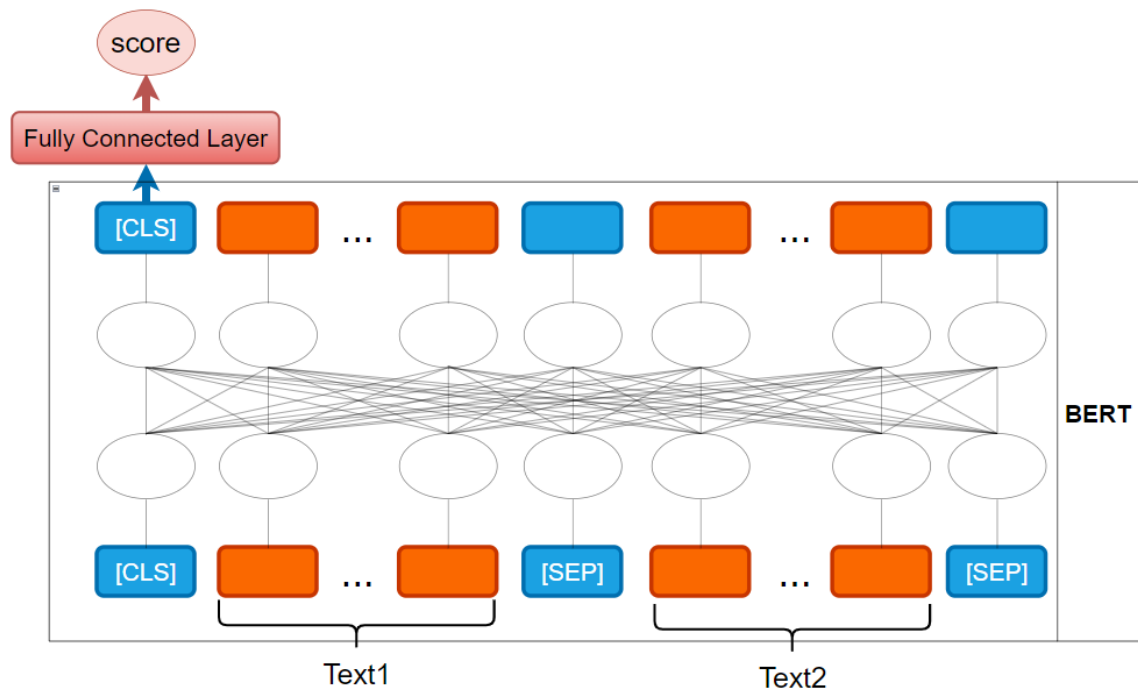
训练后的模型文件保存在/result目录下

- 预测接口：

```
python3 interface.py "text1" "text2" "saved_model"
```

预测时传入两个文本和模型路径即可得到结果

## 模型结构



## 参考文献

[1] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.