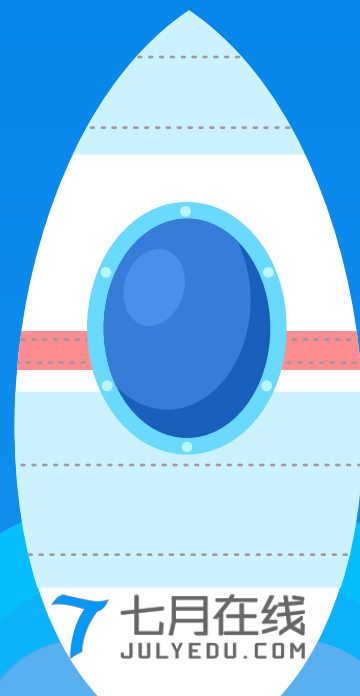


基于pytorch lightning极速完成模型搭建

Subtitle here

<https://www.julyedu.com/>

By XiangyangNi





七月在线 JULY ONLINE



公司价值观

推动社会第一
对学员负责
正直长远
持续创新改进



核心理念

培养百万AI人才 助力
AI产业发展
一站式解决人才培
养、人才招聘、业务
难题

公司简介

七月在线，成立于2015年，专注智能时代的人才培养与企业服务。

人才培养方面，讲师团队超过200人，均由大厂技术专家和国内外知名高校博士组成，总付费学员超过40万，课程数量400多门，分为初级入门课、中级集训营（定位本科和研究生的就业找工作和传统IT转型）、高级提升班（定位在职提升），且已具备完整的课程体系、服务体系、就业体系。

企业服务方面，则有以公司AI Lab领衔的近300位技术专家发力企业内训（已服务客户包括且不限于中国联通、国家电网、广发银行等等），以及公司就业部主导的给包括BAT在内的数百家企业内推合适人才。

七月在线以“培养百万AI人才 助力AI产业发展”为核心使命，致力于推动技术的传播、人才的培养、产业的发展，让全球十亿人受益。

NLP Project常需文件

- config : 设置一些基本的其他地方都要使用的全局的参数
- dataset : 设置一些关于数据处理的类
- model : 定义nlp模型的位置
- engine : 定义如何做模型的训练以及验证
- train : 整理所有的模块, 从而完成模型的finetune的过程



config.py

- 预训练模型的位置
- 超参数的设置
- 训练集，测试集的文件的位置
- 预训练完成模型保存的位置



dataset.py

- 设计dataset类对每个样本做tokenization的处理
- 将每个样本的输出做成固定的格式，从而可以成功放入模型
- 一些数据相关自定义的设计



model.py

- 定义nlp模型的地方
- 可以设计自定义的模型
- 继承nn.Module构建自定义的模型
- 直接使用Transformer的模型比如BertForSequenceClassification
- 定义简单的损失函数，对模型输出的logits进行loss的计算
- 固定模型的输出格式



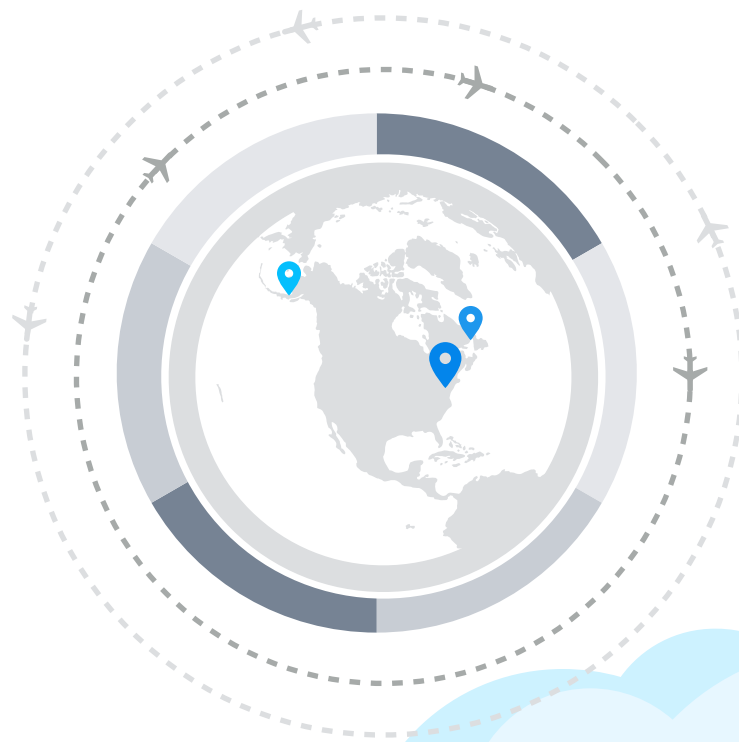
engine.py

- 通常是固定的模版
- 设计train和evaulate函数
- 如果损失函数特别的复杂，在这里定义更加方便
- 在train和evaulate中设置log日志，print出每个step和每个epoch的metrics



train.py

- 这里是将config.py, dataset.py, model.py, engine.py中所有的内容全部综合在一起
- 训练集和验证集的准备, dataloaders的准备
- 模型的准备, 优化器以及warmup的准备
- 使用train和evaluate函数做训练
- 设置提前停止, grad累计, 保存模型等动作



设计模型的缺点

- 每一个新的任务，就需要根据情景重新设计各个模块
- 每个模块中每一个细节都需要更改
- 需要自定义以及手动添加更多的代码，从而完成更加好的优化过程
- 需要通过tqdm和print的方法手动的去进行日志的记录记忆更新
- 需要通过手动记录metric的方式来记录loss和评估指标
- 评估指标没有可视化



Pytorch Lightning

- 可以解决之前所有的问题
- 专门为研究人员而开发的深度学习模型训练框架
- 由NYU和Facebook AI Research共同开发
- 只需要构建model和dataset则可以对模型做自动化的训练
- 任何自定义以及预训练的pytorch模型都能使用



Lightning的优势

- 不需要手写和维护额外代码
- 提供多种优化策略
 - 比如提前停止
 - 保存最优模型
 - 模型数据统一上传cuda
 - profiler日志的输出
 - progress_bar的输出
 - 分布式的训练
 - Tensorboard的可视化训练的中间过程
 - 等等



dataset的设计

- dataset还是原来的dataset可以直接复制过来使用
- 只是不同的数据，设计的dataset不一样而已



model的设计

- 继承LightningModule类
- 在__init__中设计好模型
- 设计forward方法
- 设计configure_optimizers方法
- 设计training_step方法
- 设计validation_step方法



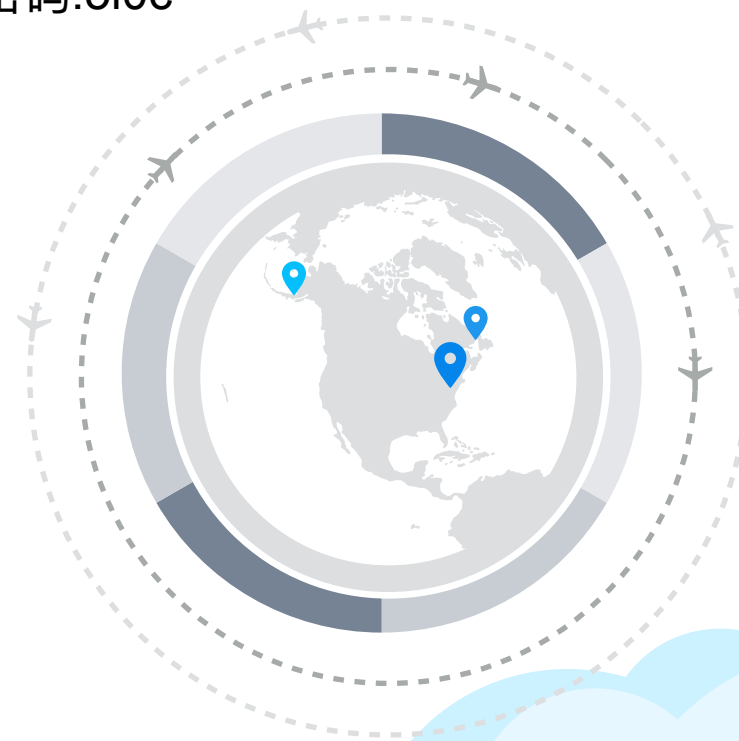
lightning中trainer类的使用

```
trainer = Trainer(  
    logger = logger,  
    min_epochs = 1,  
    max_epochs = epochs,  
    gpus = [0,1,2]  
)  
trainer.fit(  
    LModel,  
    train_dataloaders = train_dataloaders,  
    val_dataloaders = val_dataloaders  
)
```



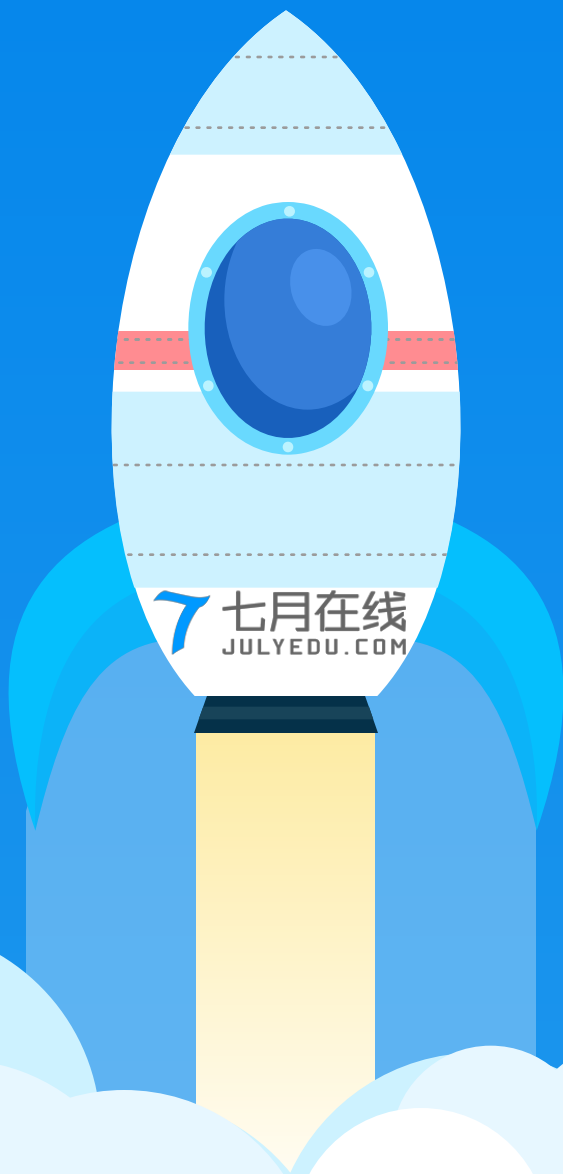
结果的展示

代码： 链接:https://pan.baidu.com/s/1Qc_b0MwRhX96A_tSyWNNtg 密码:5i0c





微信扫一扫关注我们



7 七月在线
JULYEDU.COM

THANKS

<https://www.julyedu.com>

7 七月在线
JULYEDU.COM