离线e2e模型训练代码模块接口设计说明

模型训练平台提供了训练、预测、评估的接口，可以方便模型训练人员快速熟悉训练流程，对模型训练结果进行评估。

下面将对模型训练接口进行详细说明。

# 训练接口

模型训练主要由bin/train.py作为模型训练入口，下面将对主要参数进行说明：

-train\_config ${train\_config} 模型训练配置

-data\_config ${data\_conf} 模型训练数据，conf/data.yaml

-train\_name HKUST 模型训练名字，可以根据任务指定，该参数不重要

-task\_file bin.taskegs.pytorch\_backend.task\_ctc\_att 核心任务类，无需修改

-num\_gpu 1 所使用的GPU个数

-task\_name CtcAttTask 任务名称，无需修改

-exp\_dir ${exp\_dir} 模型输出路径

-num\_epochs ${epochs} epochs个数

-seed 100 随机数种子

-checkpoint $checkpoint 要加载的基线模型

--resume\_progress 加载基线模型的训练进程

--split 对数据进行split训练，以json文件个数为准

# 预测接口

模型预测主要是对模型进行预测，由bin/predictor.py作为入口，下面对主要参数进行说明：

-train\_config ${train\_config} 模型训练配置

-data\_config ${data\_conf} 测试数据的data.yaml文件

-char\_list ${char\_list} 字典

-gpu 1 测试使用的GPU个数

-model ${checkpoint} 需要测试的模型

-exp\_dir ${exp\_dir} 测试结果文件输出路径

# 评估接口

评估接口主要是通过sclite进行打分，由bin/evaluate.py作为入口，主要参数包括：

-rec ${rec} 识别结果文件

-ref ${ref} 参考答案文件