ASRU2019中英混杂语音识别挑战赛

track3 参赛方案介绍

团队名称：WYHZ
团队单位：网易杭州研究院 语音组
团队成员：杨震 张神权 李响 刘东

2020/4/5

網易 NETEASE

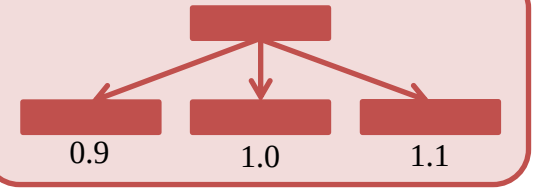
目录

- 数据预处理
- 模型训练
- 解码测试
- 分析总结

数据预处理

- 使用数据包括：
  - 官方提供 200h 中英混杂数据集
  - 官方提供 500h 纯中文数据集
  - 开源 LibriSpeech 960h 纯英文数据集

utils/perturb_data_dir_speed.sh



- 使用特征
  - 80 维 fbank 特征+pitch（steps/make_fbank_pitch.sh）

数据预处理

- 使用词典
  - 中文部分采用单字，英文部分采用BPE分词（约3000英文+7000中文）
  - 示例

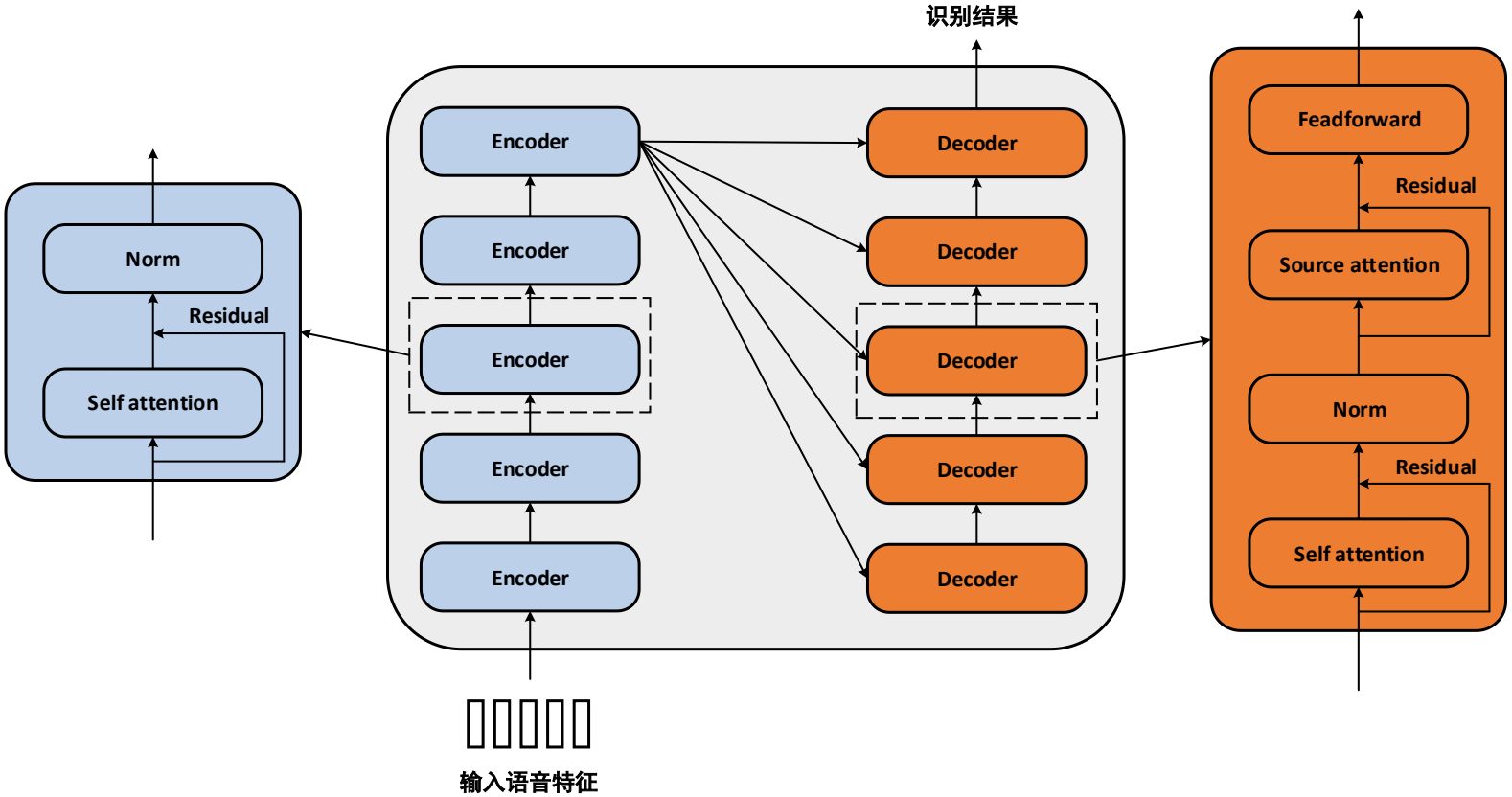
    分词前：deadline真的是第一生产力

    分词后：dead line 真 的 是 第 一 生 产 力

模型训练

- 模型信息
  - Transformer（encoder + decoder）
  - Loss：attention + ctc + language

模型训练


识别结果
Encoder
Encoder
Encoder
Encoder
Encoder
Decoder
Decoder
Decoder
Decoder
Decoder
Norm
Residual
Self attention
Feadforward
Residual
Source attention
Norm
Residual
Self attention
输入语音特征

模型训练

- 模型信息
  - Encoder layer: 6
  - Decoder layer: 12
  - Size: 2048
  - Attention dim: 512
  - Loss: attention(0.5) + ctc(0.3) + language(0.2)

模型训练

- 优化方案
  - 频谱加噪
    - 随机抹掉输入音频频谱的信息，增加模型的鲁棒性
  - 语种信息
    - 基于frame的language id，在encoder层增加CE loss
    - 基于character的language id，在decoder层增加 CE loss ***（此项最优）***
    - 上述a和b的组合，各自权重均为0.1
  - Label Smoothing
    - 缓解端到端模型存在的数据稀疏性问题

模型训练

- 训练环境
  - 数据并行的同步分布式训练
  - 大batch训练，梯度同步合并，学习率动态调整等策略
  - 16块GPU同时训练，实现快速的超参调优和模型训练及测试

解码测试

- 解码过程
  - beam search的解码方法，beam设置为10
  - one-pass解码，计算包含当前路径作为前缀的所有路径概率之和作为该步的ctc得分，融合ctc和attention得分
  - 设定阈值N，如果当前步数往前M步产生的带eos的完整路径的得分与当前最好的带eos完整路径的得分之差都小于N时，解码终止，输出当前最好的路径作为最终的识别结果（M，N=3）

参考：Hybrid CTC/Attention Architecture for End-to-End Speech Recognition

– 谢谢