Code-Switching语音识别技术方案

浙江核新同花顺网络信息股份有限公司（RoyalFlush）

目录

01

track1

指定语言模型，传统声学模型

Track1

词典生成

1）音素集：86个中文音素+CMU词典音素集
2）单词数：中文 - 86K
英文 - 94K

前端数据处理

1）自制分词工具进行抄本分析
2）数据扩充：SpecAugment
3）特征提取：音量自适应；
MFCC(40)+Pitch(3)+Ivector(100)

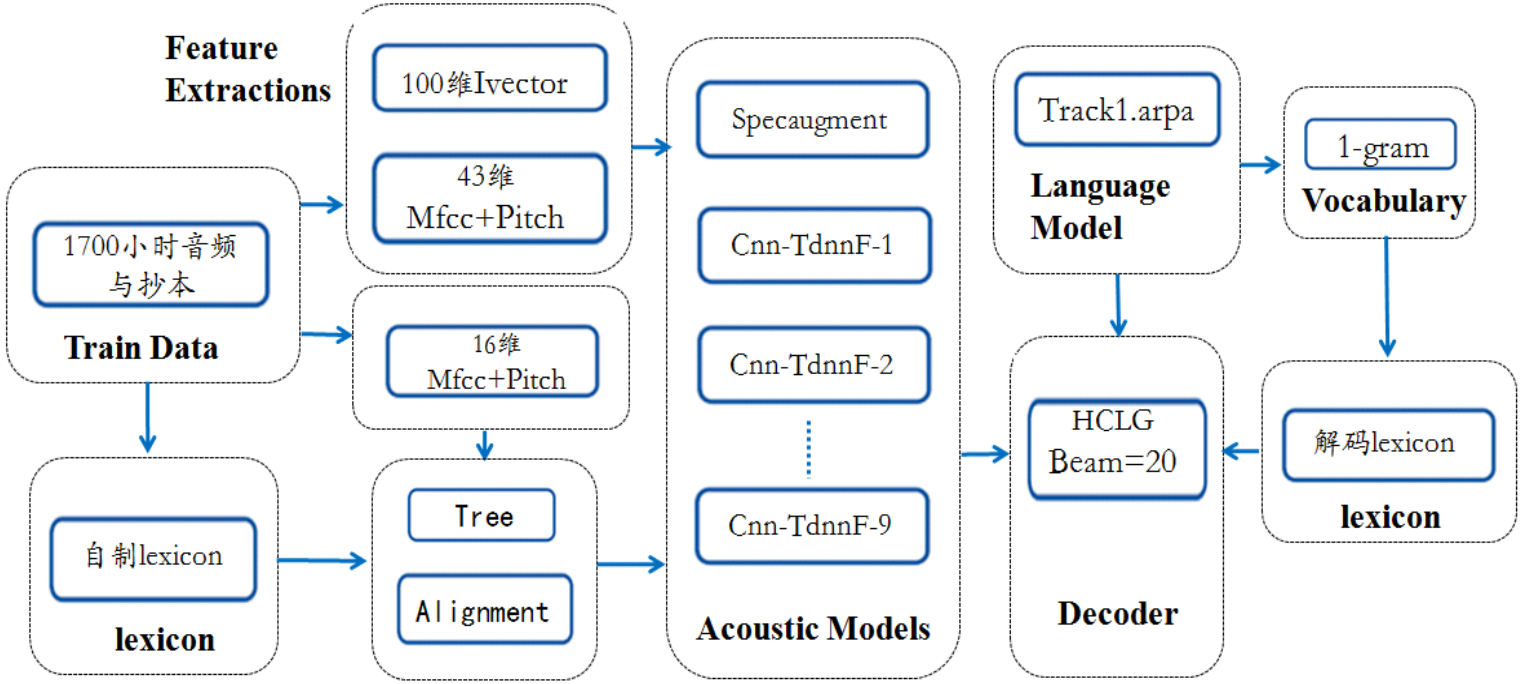

图1 track1系统流程图

Track1

其他技术

1）区分性训练
2）英文单个字母识别结果的合并处理

声学模型训练

1）模型结构：CNN+TDNNF
2）正则化技术：dropout, L2-regularization
3）学习速率：2.5e-4~2.5e-5
4）epoch：12

实验结果

表1. track1模型优化结果

优化方案	MER(%)
基础模型（baseline）	6.00
迁移学习	6.68
SMBR	**5.96**

表2 track1最终比赛测试结果

Chinese(CER%)	English(WER%)	MER(%)
5.18	18.37	6.61

02

Track 2

自制语言模型，传统声学模型

Track2

文本数据组成

1）比赛提供数据：

中文：500小时抄本20M

英文：Librispeech抄本49M

中英混合：200小时抄本7.7M

2）自有数据：

业务用纯文本和语音语料库的抄本+从Web上扒取整理的文本数据：

中文：20G

英文：8G

3）机器生成模拟文本

为了弥补中英混合文本不足，通过机器生成方式人造了一部分文本。

中英混杂：12G

Track2

文本生成策略

1）基于传统方法

对象：基础数据

方法：根据词性和词义特征进行中英文替换，同时根据词频的大小调整替换次数。

总结：

亮点-生成速度快、生成文本数量多

缺点-部分文本语句不通顺，句意不明确

} 适用于扩充基础大规模混杂文本

2）基于神经网络的文本生成

对象：dev数据

方法：利用seq2seq的Pointer-generator生成网络，对文本进行领域内的文本扩充。

总结：

亮点-运用Coverage Mechanism来解决重复生成文本的问题。

缺点-缺少标准的混杂文本作为目标输出，限制了网络模型的训练，只适用于小领域的文本扩充。

Track2

解码技术

- 大小语言模型解码，压缩HCLG大小，加速解码过程；

语言模型训练方法

1）训练工具： SRILM
2）n-gram阶数： 5-gram
3）裁剪参数： 1e-8
4）多模型插值：
   - 共分3个模型：单一英语模型、单一中文模型、中英混杂模型；
   - 根据各模型对验证集的ppl，经过compute-best-mix计算最佳混合系数；

Track2

实验结果

1）数据扩充的效果

表3.　track2语言模型用数据扩充方法的实验结果

Model	Data expansion	Vocab-Size	Dev2　(MER%)
mix-mono+code_switch.lm	No	100w	12.42
mix-code_switch.lm	Yes+通用数据扩充	100w	7.04
mix-code_switch+dev.lm	Yes+通用数据扩充+dev数据扩充	100w	5.64

2）最终提交结果

表4.　track2官方最终测试集效果

Final Model	Chinese(CER%)	English(WER%)	MER(%)
mix-code_switch+dev.lm	4.50	17.29	5.88

与Track1相比，MER降低0.8%(绝对)，12%(相对）

track3

端到端模型

Track3

前端数据处理

1）分词单元

- 汉字（中文）
- BPE subword （英文）

例文： 没 有 为 什 么 就 因 为 我 _UN LIKE

2）数据扩充

- 500小时中文+960小时英文+200x3小时中英混合(变速)

3）特征提取

- Fbank （80维）+ pitch （3维）

模型训练工具—ESPNET

1）模型结构：

Transformer

2）模型参数：

512adim-8ahead-6dlayers-2048dunits-12elayers-2048eunits

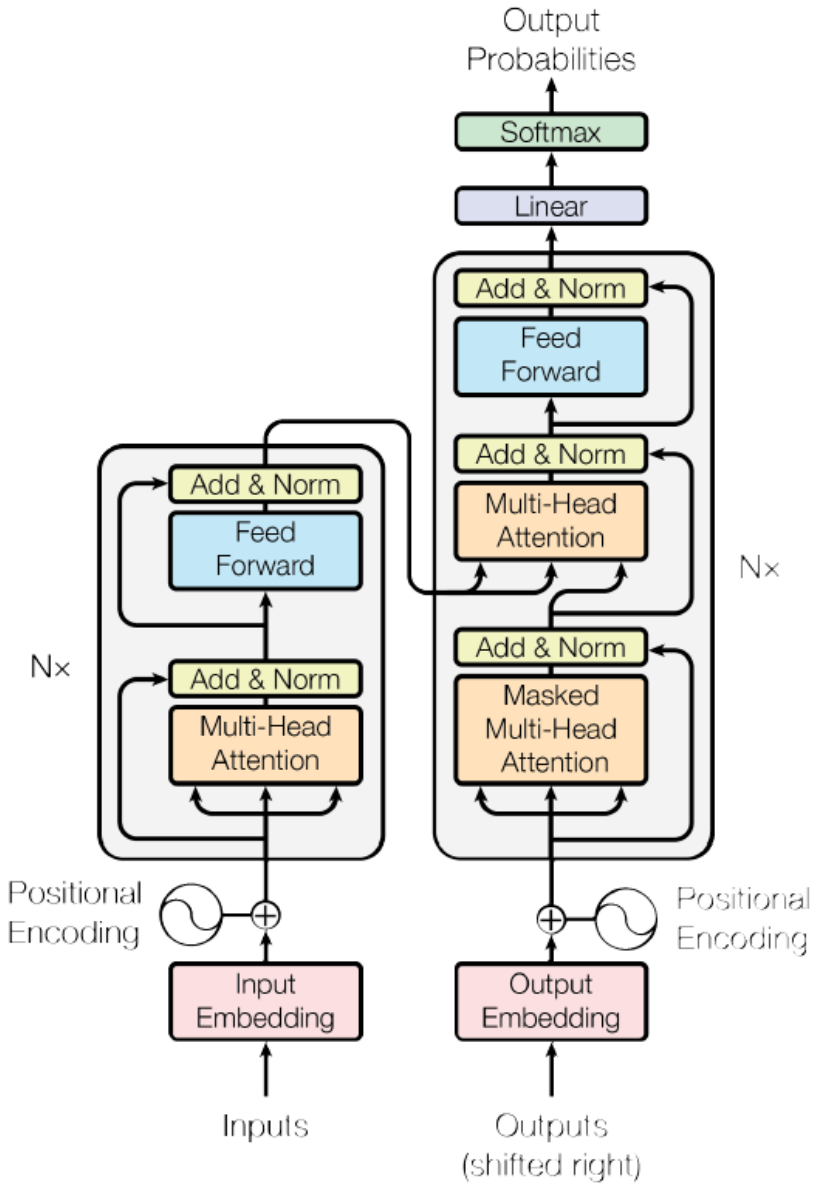

图2 Transformer结构图

Track3

■ 模型训练策略

1）CTC/Attention的混合多任务训练
2）语音识别与语种识别的联合多任务训练
3）防止过拟合策略:
(1) label smoothing (2)dropout (3) SpecAugment等

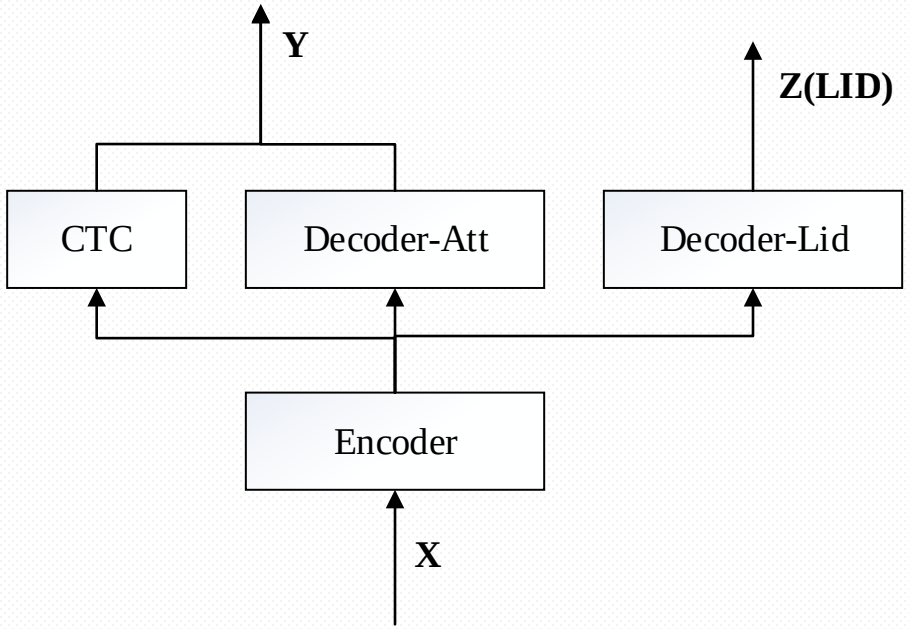

图3 模型训练结构图

■ 语言模型训练

1）训练数据：1660小时抄本
2）模型结构：4layers-2048units的LSTM

■ 后处理技术

1）英文拼写纠错模块

Track3

实验结果

1）训练方法的比较

表5 track3中模型架构，数据变速增强及后处理的实验结果

Model	Train Data(hours)	Spell	Dev2(%)
Att+CTC	1660	No	12.80
Att+CTC+LID	1660	No	12.48
Att+CTC+LID	2060	No	10.04
Att+CTC+LID	2060	Yes	10.00

① 联合训练有一定效果
② 数据扩充效果较好
③ 拼写纠错可以改善对英文的识别率，但是对整体的识别率效果提升不明显。

2）最终提交结果

表6 track3最终测试结果

Final Model	Chinese (CER%)	English (WER%)	MER (%)
Att+CTC+LID	7.49	21.40	9.00

谢 谢