

LOGFILE

一、测试 **Audioseg** 和 **DarITK** 工具包。

二、系统说明

1 工具包

Audioseg

```
/home/users/wangjun/audioseg
bin/          #编译生成的可执行文件
include/     #.h 文件  log/          share/
test/        #提供的测试 shell, 测试单条指令功能
data/        #数据目录
lib/         #生成的静态库
res/         #输出目录
src/         #source code
test.sh      #编写的 shell 脚本, 执行 darization 任务
```

DarITk

```
/home/users/wangjun/darizationTools/spkr_diar_ib_rel/
bin/         #编译生成的可执行文件目录 aibfeat 提 ib 特征
              realign 执行 realign
              realign_simple_out 功能同上, 修改了代码中写输出文件格式
config/ib.cfg #配置文件
mat/         #mat 代码, 执行 ib 聚类
src/         #source code
tmp/         #执行的中间过程文件和结果输出文件目录, *.rttm 最终文件 *.scp
分割文件
data/        #IS1004c.rttm IS1005c.rttm 工具包提供的 groundtruth
              #IS1004c.scp IS1005c.scp 工具包提供的分割文件, 用于 aibfeat 输入
              #md-eval-v21.pl score.sh 打分工具
              #子目录 mfcc ms del fdlp 相应特征目录
scripts/     #编写的 shell 脚本, 执行测试任务
```

2 测试数据

NIST 数据和手动拼接语音

```
/home/users/wangjun/data/
hyp-groundtruth/ #groundtruth
hypreseg/       #Nick 提供的经过 ALIZE 系统的分割结果
mfcc/          #自己提取的特征目录
sphere/        #NIST 的.sph 文件, 8k 16 位双声道 PCM16
              #2004.wav 和 2005.wav 为 cooledit 转换后单声道的 wav
              #mfss.sh Config.txt Filelist.txt 为 HTK 提特征工具及配置文件
              #1.wav 2.wav.3.wav 为手动拼接文件, 除了 1.wav 需要重新记录拼接顺序,
手动标记 groundtruth
```

DariTK 工具包自带特征文件

```
/home/users/wangjun/darizationTools/spkr_diar_ib_rel/data/  
#IS1004c.rttm IS1005c.rttm 工具包提供的 groundtruth  
#IS1004c.scp IS1005c.scp 工具包提供的分割文件，用于 aibfeat 输入  
#/home/users/wangjun/darizationTools/spkr_diar_ib_rel/README 提供了结果
```

3 特征工具

注：工具包均不提供特征提取工具，Audioseg 推荐安装了 spro5.0，DariTK 查看了代码好像必须使用 HTK 提取，并用 spro 格式的 mfcc 测试不通过。估计是读取的文件头不同。

Spro5.0

```
/home/users/wangjun/spro-5.0/bin/bin  
mfcc.sh #编写的提特征脚本
```

HTK HCopy

```
/home/users/wangjun/data/  
sphere/ #mfss.sh Config.txt Filelist.txt 为 HTK 提特征工具及配置文件
```

4 其它目录说明

gsl

```
/home/users/wangjun/gsl #gsl 下载目录  
/home/users/wangjun/gsl-1.15 #gsl 安装目录
```

ALIZE

```
/home/users/wangjun/seg #Nick 提供的包，已经编译通过，脚本及配置还在调整  
/home/users/wangjun/ALIZE #自己下载安装的包，已经编译通过
```

Shout #另一个 diarization 工具，为语音识别前端，未测试

```
/home/users/wangjun/darizationTools/shout
```

5 打分工具

```
NIST 下载的打分工具  
/nfs/disk/work/users/wangjun/seg/upload/seg_scr.v21/seg_scr.v21
```

同上，Fabio 用于系统打分

```
/home/users/wangjun/darizationTools/spkr_diar_ib_rel/data  
#md-eval-v21.pl score.sh 打分工具
```

三、实验

1 分别用 spro 和 htk 提特征,特征均为加一阶 delta 加 energy。

```
spro 特征后缀 .spro.mfcc  
htk 特征后缀 .htk.mfcc
```

2 测试 audioseg

注: **Audioseg** 输入特征要求是 **spro**, 尝试用 **htk** 无法运行。

ssad 后缀 .sad

sbic 后缀 .bic

scluster 后缀 .clu

Sat Nov 24 20:07:57 CST 2012

Sat Nov 24 20:07:57 CST 2012

Sat Nov 24 20:08:02 CST 2012

Sat Nov 24 20:11:17 CST 2012

Sat Nov 24 20:11:21 CST 2012

Sat Nov 24 20:14:36 CST 2012

Sat Nov 24 20:14:40 CST 2012

Ok

Nick 提供的数据不做 sad, 用全语音段直接做 bic。

res1/2004.bic 2004.seg o2004.clu

2005.bic 2005.seg o2005.clu

3 测试 DariTK

不同 scp 文件的结果存放不同目录

整段语音不分割的 scp 结果存放于 tmpres

用 tmpres 输出的 scp 作为输入结果存放于 tmpres2

用 sad 结果作为 scp 输入结果存放于 tmpres3

daritk 提供的 scp 和特征输入, 结果存放于 fabiotmp

daritk 提供的特征和整段不分割语音 scp 作为输入, 结果存放于 fabiotmpuniscp

4 打分

82.86

69.58

Audioseg				
DER(SPEAKER ERROR TIME)	对角阵	f	i	fi
1.wav	10.28	61.98 (?)	10.28	61.98
2004.wav	60.03	27.01	60.03	27.01

2005.wav	82.04	40.06	82.04	40.06
Daritk				
DER(SPEAKER ERROR TIME)	整段语音 scp	等间隔 scp(2)	Fabio 提供的 scp	sad 的 scp
1.wav				
2004.wav	47.01 (31.6)	47.01 (31.6)		27.99 (24.5)
2005.wav	37.84 (20.9)	37.84 (20.9)		23.94 (8.3)
IS1004c	44.43		15.80(11.2)	用 fabio 提供 scp 输出的 tmp 中 scp 作为输入 16.04
IS1005c			18.69(15.8)	
Nick 提供的测试数据				
2004		4.14 (4.14)		
2005		2.50 (2.50)		

5 结论

- (1) audioseg 使用 bic 检测 speaker turn 后做 cluster, 对于较短语音且说话人转换分割点明显的语音, 能得到较好效果。(不明白为何采用 full matrix 结果很差?)
- (2) audioseg 对于较长语音且分割转换较为频繁的语音, bic 的结果首先就不准确, 采用 full matrix 运算等得到结果。
- (3) 参数设置对 audioseg 影响很大, 尝试了不同 sad 能量域值, 最短静音长度, bic 的参数等。对结果影响很大, 尚没试出较好的参数。目前的 FAILSE ALARM ERROR 很大, 应该和跟 sad 和 bic 参数相关。
- (4) Audioseg 应该可以实现 bottomup 系统并 alignment 的步骤, 代码上需要调整。
- (5) DatiTK 对于输入的 scp 文件依赖度较高, 采用 vad 分割后的语音能得到较好效果。
- (6) DariTK 运行速度更快。
- (7) Nick 提供的数据只有 speaker error 一项, 即 () 中内容。感觉和 fabio 提供的 scp 是一个道理。Fabio 提供的 ground truth 和 scp 是一样的分割点, 然后做聚类。同样, 如果用 nick 提供的 groundtruth 文件生成 scp 文件的分割点, 一定会得到很好的效果。