

PocketSphinx 语音识别系统语言模型的训练和声学模型的改进

1: 训练语言模型:

需要文件: 写有包含中文内容的语料库, 网上工具: <http://www.speech.cs.cmu.edu/tools/lmtool-new.html>, 上传中文内容的语料库的文件, 就可以返回语言模型文件。如果你的语言模型比较大的话, 最好就转换为 CMU 的二进制格式 (DMP), 这样可以加快加载语言模型的速度, 减少解码器初始化的时间。但对于小模型来说, 就没有这个必要。面令:

```
$sphinx_lm_convert -i 要转换的模型 -o 输出的模型
```

2.声学模型改进:

需要的文件: sound.txt : 记录音频文件的地址, 如 (下面例子表明音频文件是当目录下的 001101_L~001105_L) :

```
001101_L
001102_L
001103_L
001104_L
001105_L
```

direction.txt: 中文句子与语音文件的对应, 如:

```
<s> 中关村</s> (001101_L)
<s> 北京人</s> (001102_L)
<s> 生活</s> (001103_L)
<s> 朋友</s> (001104_L)
<s> 熬夜</s> (001105_L)
```

zh.dic : 字典文件,可以由原有模型得到, 也可以自己新建, 如:

```
中关村 zh ong g uan c un
北京人 b ei j ing r en
生活 sh eng h uo
朋友 p eng y ou
熬夜 ao y ie
```

音频文件: 地址与 sound.txt 里的一致。

已训练好的声学模型 (zh)

2.1: 提取特征:

```
$sphinx_fe -argfile zh/feat.params -samprate 16000 -c sound.txt -di . -do . -ei wav -eo mfc -mswav yes
```

2.2: 转化 sendump 和 mdef 文件

确保 mixture_weights 文件在的模型里, 没有在 <https://sourceforge.net/projects/cmuspinx/files/> 下载, 然后我们需要把声学模型的 mdef 文件转换为 SphinxTrain 训练工具可以使用的 plain text 格式

```
$pocketsphinx_mdef_convert -text zh/mdef zh/mdef.txt
```

2.3: 累加观察序列

把 bw 和 map_adapt 和 mk_s2sendump 和 mllr_solve 程序都拷贝到当前目录路径：三个文件的路径一般是：/usr/local/libexec/sphinxtrain。然后开始统计：

```
./bw -hmmidir zh -moddefn zh/mdef.txt -ts2cbfn .ptm. -feat s2_4x -cmn current -agc none -dictfn test.txt -ctlfm suond.txt -lsnfn direction.txt -accumdir .
```

（面令参数与模型内 feat.params 文件一致）

2.4、创建 MLLR（最大似然线性回归算法）变换：

```
./mllr_solve -meanfn zh/means -varfn zh/variances -outmllrfn mllr_matrix -accumdir .
```

2.5:通过 MAP 更新声学模型

```
$cp zh/ -rf zhadapt
```

```
./map_adapt -meanfn zh/means -varfn zh/variances -mixwfn zh/mixture_weights -tmatfn zh/transition_matrices -accumdir . -mapmeanfn zhadapt/means -mapvarfn zhadapt/variances -mapmixwfn zh/mixture_weights -maptmfn zhadapt/transition_matrices
```

2.6、重新创建适应的 sendump 文件

通过已经更新的 mixture_weights 文件来重新创建 sendump 文件，更新完可以将 mixture_weights 删除以节省空间：

```
./mk_s2sendump -pocketsphinx yes -moddefn zhadapt/mdef.txt -mixwfn zhadapt/mixture_weights -sendumpfn zhadapt/sendump
```

现在就已经已经拥有一个改进了的适应你的语音的声学模型了。新的声学模型在 zhadapt 目录下。

3.测试模型：

测试数据：分别包含中文语音 “中关村”、“朋友”、“熬夜”，的音频文件。

命令：#pocketsphinx_continuous -hmm 声学模型文件路径 -lm 语言模型文件名 -dict 字典文件

-infile <file.wav>。如：\$ pocketsphinx_continuous -hmm zhadapt -lm 6252.lm -dict test.txt -infile 4.wav

测试结果：

输入：“中关村”的音频文件

输出：

```
whang@whang: ~/yuyinshibie/newmodel
文件(F) 编辑(E) 查看(V) 搜索(S) 终端(T) 帮助(H)
INFO: ngram_search_fwdflat.c(302): Utterance vocabulary contains 6 words
INFO: ngram_search_fwdflat.c(948):      484 words recognized (4/fr)
INFO: ngram_search_fwdflat.c(950):      17026 senones evaluated (123/fr)
INFO: ngram_search_fwdflat.c(952):      6686 channels searched (48/fr)
INFO: ngram_search_fwdflat.c(954):      1317 words searched (9/fr)
INFO: ngram_search_fwdflat.c(957):      288 word transitions (2/fr)
INFO: ngram_search_fwdflat.c(960): fwdflat 0.08 CPU 0.061 xRT
INFO: ngram_search_fwdflat.c(963): fwdflat 0.08 wall 0.061 xRT
INFO: ngram_search.c(1250): lattice start node <s>.0 end node </s>.100
INFO: ngram_search.c(1276): Eliminated 2 nodes before end node
INFO: ngram_search.c(1381): Lattice has 88 nodes, 108 links
INFO: ps_lattice.c(1380): Bestpath score: -4429
INFO: ps_lattice.c(1384): Normalizer P(0) = alpha(</s>:100:136) = -295066
INFO: ps_lattice.c(1441): Joint P(0,S) = -311449 P(S|0) = -16383
INFO: ngram_search.c(872): bestpath 0.00 CPU 0.000 xRT
INFO: ngram_search.c(875): bestpath 0.00 wall 0.000 xRT
中关村
INFO: ngram_search_fwdtree.c(429): TOTAL fwdtree 0.35 CPU 0.111 xRT
INFO: ngram_search_fwdtree.c(432): TOTAL fwdtree 0.35 wall 0.111 xRT
INFO: ngram_search_fwdflat.c(176): TOTAL fwdflat 0.10 CPU 0.034 xRT
INFO: ngram_search_fwdflat.c(179): TOTAL fwdflat 0.10 wall 0.034 xRT
INFO: ngram_search.c(303): TOTAL bestpath 0.00 CPU 0.000 xRT
INFO: ngram_search.c(306): TOTAL bestpath 0.00 wall 0.000 xRT
whang@whang: ~/yuyinshibie/newmodel$
```

输入：“朋友”的音频文件

输出:

```
whang@whang: ~/yuyinshibie/newmodel
文件(F) 编辑(E) 查看(V) 搜索(S) 终端(T) 帮助(H)
INFO: ngram_search_fwdfat.c(302): Utterance vocabulary contains 4 words
INFO: ngram_search_fwdfat.c(948):      585 words recognized (5/fr)
INFO: ngram_search_fwdfat.c(950):      13345 senones evaluated (108/fr)
INFO: ngram_search_fwdfat.c(952):      5188 channels searched (42/fr)
INFO: ngram_search_fwdfat.c(954):      1141 words searched (9/fr)
INFO: ngram_search_fwdfat.c(957):      176 word transitions (1/fr)
INFO: ngram_search_fwdfat.c(960): fwdfat 0.06 CPU 0.051 xRT
INFO: ngram_search_fwdfat.c(963): fwdfat 0.06 wall 0.051 xRT
INFO: ngram_search.c(1250): lattice start node <s>.0 end node </s>.78
INFO: ngram_search.c(1276): Eliminated 3 nodes before end node
INFO: ngram_search.c(1381): Lattice has 138 nodes, 143 links
INFO: ps_lattice.c(1380): Bestpath score: -2990
INFO: ps_lattice.c(1384): Normalizer P(0) = alpha(</s>:78:121) = -210856
INFO: ps_lattice.c(1441): Joint P(0,S) = -226712 P(S|0) = -15856
INFO: ngram_search.c(872): bestpath 0.00 CPU 0.000 xRT
INFO: ngram_search.c(875): bestpath 0.00 wall 0.000 xRT
朋友
INFO: ngram_search_fwdtree.c(429): TOTAL fwdtree 0.34 CPU 0.113 xRT
INFO: ngram_search_fwdtree.c(432): TOTAL fwdtree 0.34 wall 0.113 xRT
INFO: ngram_search_fwdfat.c(176): TOTAL fwdfat 0.08 CPU 0.028 xRT
INFO: ngram_search_fwdfat.c(179): TOTAL fwdfat 0.08 wall 0.028 xRT
INFO: ngram_search.c(303): TOTAL bestpath 0.00 CPU 0.000 xRT
INFO: ngram_search.c(306): TOTAL bestpath 0.00 wall 0.000 xRT
whang@whang:~/yuyinshibie/newmodel$
```

输入: “熬夜” 的音频文件

输出:

```
whang@whang: ~/yuyinshibie/newmodel
文件(F) 编辑(E) 查看(V) 搜索(S) 终端(T) 帮助(H)
INFO: ngram_search_fwdfat.c(302): Utterance vocabulary contains 2 words
INFO: ngram_search_fwdfat.c(948):      1396 words recognized (4/fr)
INFO: ngram_search_fwdfat.c(950):      5607 senones evaluated (18/fr)
INFO: ngram_search_fwdfat.c(952):      2491 channels searched (7/fr)
INFO: ngram_search_fwdfat.c(954):      2491 words searched (7/fr)
INFO: ngram_search_fwdfat.c(957):      76 word transitions (0/fr)
INFO: ngram_search_fwdfat.c(960): fwdfat 0.04 CPU 0.011 xRT
INFO: ngram_search_fwdfat.c(963): fwdfat 0.04 wall 0.012 xRT
INFO: ngram_search.c(1250): lattice start node <s>.0 end node </s>.281
INFO: ngram_search.c(1276): Eliminated 0 nodes before end node
INFO: ngram_search.c(1381): Lattice has 361 nodes, 576 links
INFO: ps_lattice.c(1380): Bestpath score: -2108
INFO: ps_lattice.c(1384): Normalizer P(0) = alpha(</s>:281:313) = -125377
INFO: ps_lattice.c(1441): Joint P(0,S) = -144417 P(S|0) = -19040
INFO: ngram_search.c(872): bestpath 0.00 CPU 0.000 xRT
INFO: ngram_search.c(875): bestpath 0.00 wall 0.000 xRT

INFO: ngram_search_fwdtree.c(429): TOTAL fwdtree 0.32 CPU 0.101 xRT
INFO: ngram_search_fwdtree.c(432): TOTAL fwdtree 0.32 wall 0.101 xRT
INFO: ngram_search_fwdfat.c(176): TOTAL fwdfat 0.04 CPU 0.012 xRT
INFO: ngram_search_fwdfat.c(179): TOTAL fwdfat 0.04 wall 0.012 xRT
INFO: ngram_search.c(303): TOTAL bestpath 0.00 CPU 0.000 xRT
INFO: ngram_search.c(306): TOTAL bestpath 0.00 wall 0.000 xRT
whang@whang:~/yuyinshibie/newmodel$
```

没有输出