基于差分特征和高斯混合模型的湖南方言识别

王岐学,钱盛友,赵新民
WANG Qi-xue,QIAN Sheng-you,ZHAO Xin-min

湖南师范大学 物理与信息科学学院,长沙 410081
College of Physics and Information Science,Hunan Normal University,Changsha 410081,China

WANG Qi-xue,QIAN Sheng-you,ZHAO Xin-min.Hunan dialects identification based on GMM and difference speech feature.Computer Engineering and Applications,2009,45(35):129-131.

Abstract: Rhythm of speech is an important acoustic distinction between different Chinese dialects,and the difference feature is an important reflection of rhythm.While difference features $\Delta MFCC$ & $\Delta\Delta MFCC$ are used as characteristic parameters and Gaussian Mixture Model(GMM) is used as a trained model,the dialect can be identified through calculating the likelihood probability of the test samples.Changsha dialect,Shaoyang dialect,Hengyang dialect and Mandarin have been investigated with this method, and its effect has been compared with the effect using *MFFC* as characteristic parameters.Experiment results show that a more high recognition rate and better anti-noise performance can be obtained by GMM trained with difference feature.

Key words: differential feature;Gaussian Mixture Model(GMM);dialects identification

摘 要:语音的韵律是区分汉语方言的重要语音声学特征,而语音的差分特征是语音韵律的重要体现。采用差分特征 $\Delta MFCC$ 和 $\Delta\Delta MFCC$ 作为特征参数,用高斯混合模型(GMM)作为训练模型,通过计算测试样本的似然概率来识别方言的类型。用该方法对长沙方言、邵阳方言、衡阳方言和普通话进行了识别研究,并与采用 *MFCC* 作为特征参数的识别效果进行了比较。实验结果表明差分特征具有识别率高、抗噪声性能更好等优点。

关键词:差分特征;高斯混合模型;方言识别
DOI:10.3778/j.issn.1002-8331.2009.35.039 **文章编号**:1002-8331(2009)35-0129-03 **文献标识码**:A **中图分类号**:TP391

1 引言

汉语方言的声学特征差异是由多方面的原因产生的,但最主要表现在声韵母系统和语音韵律的不同。声韵母的不同主要表现在元音中,表现形式为共振峰上的差异[1]。语音韵律的不同主要表现为基音音调和语速之间的差别,其中用于获取语音特征向量连续动态变化轨迹的差分特征表现最为明显,该特征向量的变化体现了说话人言语和韵律的变化。一般而言,由言语变化带来的特征向量的变化速度较快,而韵律的变化则是一个渐进的过程,体现了发音的高低起落。方言发音的习惯差异主要表现在语音频谱结构的时间变化上,即特征参数的动态特征[2]。这类特性相对稳定且比较容易模仿,代表性的有 Δ 倒谱和 $\Delta\Delta$ 倒谱。倒谱参数所包含的信息较其他参数多,它是目前普遍采用的语音特征参数。特征参数的好坏直接影响方言识别系统的性能。借助倒谱系数和它的差分系数对长沙方言、邵阳方言、衡阳方言以及普通话进行基于高斯混合模型的方言识别建模,并对差分特征的优异性进行了分析。

2 特征参数的提取

倒谱系数反映了声道的共振性能,常用的倒谱系数有:线性预测倒谱系数(LPCC)和 Mel 倒谱系数(MFCC)。Mel 倒谱系数利用了听觉原理和倒谱的解相关特性,另外 MEL 倒谱对卷积性信道失真有补偿的能力[3]。基于这些原因,Mel 倒谱系数被证明是语音识别中最成功的特征描述之一,采用 MFCC 及其动态特征一阶、二阶差分作为方言识别的特征参数。

特征参数(*MFCC*、$\Delta MFCC$ 和 $\Delta\Delta MFCC$)的算法流程如图 1 所示。首先对语音信号 $S(n)$进行预加重、加 hamming 窗分帧等处理,其中帧长为 240(30 ms,*fs* 为 8 kHz);然后对每帧语音信号 $x(n)$进行 DFT/FFT 变换得到信号的频谱 $X(k)$;用 M 阶 Mel 滤波器对得到的每帧信号的离散功率谱($|X(k)|^2$)进行滤波(传递函数为 $H_m(k)$),并求取相应的对数能量谱。其公式为:

$$s(m)=\ln\left(\sum_{k=0}^{N-1}|X(k)|^2 H_m(k)\right) \quad (1)$$

将对数能量谱 $s(m)$经过 DCT 后,求得 MFCC,公式为:

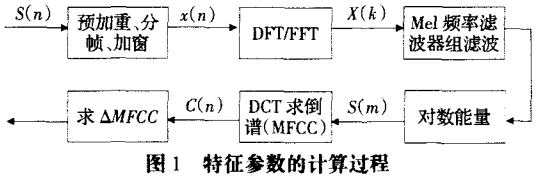

图 1 特征参数的计算过程

作者简介:王岐学(1985-),男,硕士研究生,研究方向语音信号处理;通讯作者:钱盛友(1965-),男,教授,博导,主要研究方向为信号检测与处理及智能仪器等。

收稿日期:2009-07-28 **修回日期**:2009-10-19

$$C(n)=w_m\sum_{m=1}^{M}s(m)\cos\left(\frac{\pi n(2m+1)}{2M}\right),0\leqslant m\leqslant M,n=1,2,\cdots,K \quad (2)$$

其中 w_m 是倒谱提升窗口。

提取了 MFCC 参数后，再求差分特征参数 $\Delta MFCC$ 和 $\Delta\Delta MFCC$，计算公式为：

$$\Delta C_n(t)=\frac{\partial C_n(t)}{\partial t},\Delta\Delta C_n(t)=\frac{\partial^2 C_n(t)}{\partial t^2} \quad (3)$$

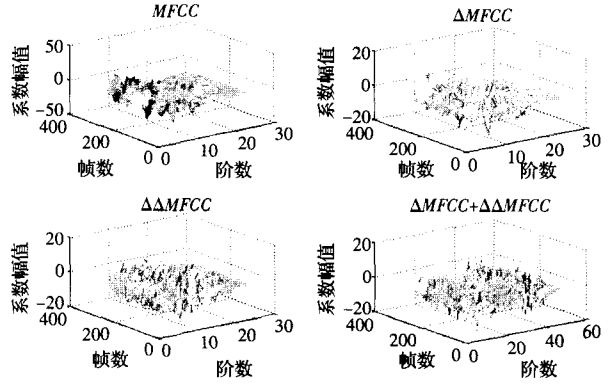

图 2 一段语音的 $MFCC$、$\Delta MFCC$、$\Delta\Delta MFCC$、$\Delta MFCC+\Delta\Delta MFCC$

3 模型及参数估计方法

3.1 高斯混合模型(GMM)

GMM 是用多个 N 维高斯分布概率密度函数的加权组合来描述矢量在概率空间分布的混合模型[4]。一个 M 阶高斯混模型就是 M 个单高斯分布的加权组合。一个完整的 GMM 可表示为 $\Theta=\{a_i,\vartheta_i;i=1,2,\cdots M\}$，其中 a_i 是 M 个单高斯分布的第 i 个分布的权值，满足 $\sum_{i=1}^{M}a_i=1$，其高斯分布的参数是 $\theta_i=\{u_i,\Sigma_i\}$；对应的概率密度函数是：

$$f_i(\boldsymbol{X}|\theta_i)=2\pi^{-\frac{n}{2}}|\Sigma_i|^{-\frac{1}{2}}\exp[-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{u}_i)^{\mathrm{T}}\Sigma_i^{-1}(\boldsymbol{X}-\boldsymbol{u}_i)] \quad (4)$$

其中 $\boldsymbol{X}$ 是 n 维高斯变量，u_i 是矢量 $\boldsymbol{X}$ 的第 i 个模型均值(n 维)，方差为 Σ_i。由 M 个模型产生它的概率密度函数为：

$$f(X|\Theta)=\sum_{i=1}^{M}a_i f_i(X|\theta_i) \quad (5)$$

3.2 参数估计

模型的参数估计就是根据样本观测序列推断出模型的参数，包括权值 a_i、各单高斯分布均值 u_i 和方差 Σ_i，也叫模型的训练。由于已经确定了模型的概率分布形式，所以采用最大似然估计的方法计算[5-7]。

设已知 d 个样本序列 $x_1,x_2,\cdots x_d$，则其似然函数为：

$$L_d(\Theta)=f(x_1,x_2,\cdots x_d,\Theta)=\prod_{i=1}^{d}f(x_i,\Theta)=\prod_{i=1}^{d}(\sum_{j=1}^{M}a_j f_j(x_i|\theta_j)) \quad (6)$$

能使概率密度函数达到最大的哪些值，就是它的相应的随机变量最大可能值，因此最大似然估计是使似然函数达到最大值的一种最佳估计。又因为似然函数是单调的，$\ln L$ 与 L 同时达到最大值，因此可以通过 $\ln L$ 来求最大值。其中：

$$\ln L_d=\sum_{i=1}^{d}\ln f(x_i,\Theta)=\sum_{i=1}^{d}\ln[\sum_{j=1}^{M}a_j f_j(x_i|\theta_j)] \quad (7)$$

采用 EM 算法来对上式进行简化，EM 算法是一种缺失数据情况下参数的估计算法。所谓数据不完备有两种情况：一种是观察的过程中本身的限制；另一种是似然函数直接优化十分困难，要引入额外的参数进行优化。这就定义了原始数据加上额外数据组成“完备数据”，原始观察数据就成了“不完备数据”。设观察到的数据是 X，完备数据就是 $Y=(X,Z)$，Z 是缺失的数据。相应的对数似然函数是：

$$\ln L_d(\Theta|X,Z)=\sum_{i=1}^{d}\ln f(x_i,\Theta)=\sum_{i=1}^{d}\ln[\sum_{j=1}^{M}a_j f_j(x_i|\theta_j)] \quad (8)$$

4 模型训练

实验中对长沙方言、邵阳方言、衡阳方言、普通话等分别选取不同年龄和性别的 12 个说话人，每人随机读 240 句左右的语音样本并提取其特征参数对混合模型进行训练后得到的 4 个高斯混合模型$\{\Theta_i\}_{i=1}^{4}$，分别代表 4 种语音的识别模型。图 3 是邵阳方言训练 16 阶高斯混合模型后得到的模型各单高斯分布的权值。

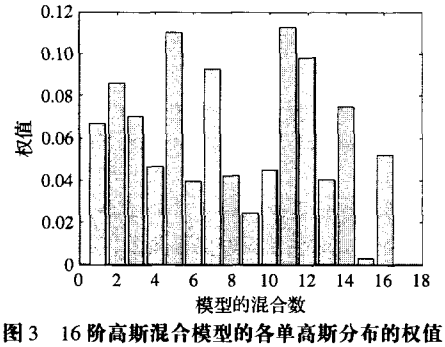

图 3 16 阶高斯混合模型的各单高斯分布的权值

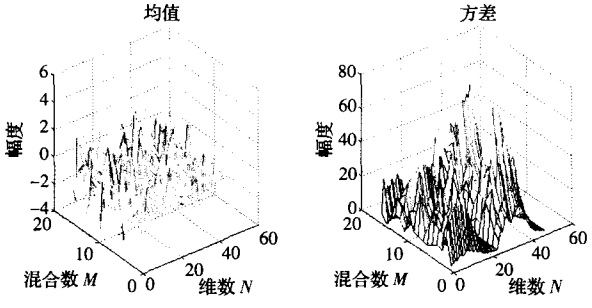

图 4 高斯分布均值 u_i 及方差 $\sum_i$

5 实验结果及分析

对测试语言进行特征提取，并用方言模型计算其似然概率值。在方言鉴别中，似然概率最大的就是测试语音的识别结果。计算公式为：

$$\hat{k}=\arg\max_{i=1}^{4}\sum_{j=1}^{d}\ln f(x_j,\Theta_i) \quad (9)$$

分别用各个模型对测试语句进行测试，计算其相应的似然概率。其中 GMM 模型的混合数为 16，采用的特征参数为 48 维的 $\Delta MFCC+\Delta\Delta MFCC$。为研究差分特征的识别性能，在不同噪声环境下用上述方法对长沙话、邵阳话和衡阳话等湖南方言及普通话进行了测试，并与用 24 维 MFCC 作为特征参数的测试结果进行了比较。测试样本为每种语音取 10 s 左右，然后加 10 dB、30 dB 的白噪声，分别用 4 种模型计算其似然概率，结果如表 1 所示。

表 1 不同特征参数及不同信噪比条件下各方言的似然概率

测试方言(SNR)		模型及特征参数选取							
		长沙		邵阳		衡阳		普通话	
		MFCC	*ΔMFCC+ΔΔMFCC*	*MFCC*	*ΔMFCC+ΔΔMFCC*	*MFCC*	*ΔMFCC+ΔΔMFCC*	*MFCC*	*ΔMFCC+ΔΔMFCC*
长沙	0 dB	**0.60**	**0.71**	0.27	0.17	0.30	0.18	0.24	0.10
	10 dB	**0.56**	**0.68**	0.23	0.19	0.29	0.17	0.23	0.15
	30 dB	**0.49**	**0.68**	0.26	0.16	0.23	0.11	0.27	0.13
邵阳	0 dB	0.30	0.16	**0.63**	**0.74**	0.11	0.18	0.33	0.11
	10 dB	0.21	0.19	**0.54**	**0.73**	0.29	0.15	0.31	0.17
	30 dB	0.23	0.14	**0.49**	**0.71**	0.26	0.18	0.23	0.14
衡阳	0 dB	0.29	0.19	0.31	0.17	**0.55**	**0.70**	0.26	0.13
	10 dB	0.21	0.13	0.23	0.19	**0.50**	**0.51**	0.25	0.19
	30 dB	0.20	0.15	0.26	0.12	**0.47**	**0.68**	0.28	0.16
普通话	0 dB	0.27	0.13	0.26	0.11	0.26	0.14	**0.63**	**0.77**
	10 dB	0.26	0.17	0.25	0.14	0.30	0.17	**0.58**	**0.78**
	30 dB	0.30	0.15	0.19	0.12	0.27	0.19	**0.52**	**0.74**

6 结论

仿真结果表明无论在有无噪声的情况下,采用差分特征参数作为识别参数时,正确识别的似然概率远比传统 *MFCC* 作为识别参数时要大,而错误识别时对应的似然概率也要小,所以差分特征比传统特征有更好的性能。在噪声环境下,差分特征的似然概率变化比较小,可见差分参数具有较强的抗噪声能力。实验还发现似然概率有时会出现突变,与理论值有偏差,其识别率不是最佳。系统实际上是基于统计模式的识别方式,其结果依赖于统计样本数,样本不充分,统计结果受到一定的影响。另外,测试语句可能并不具有代表性,相应语句两种方言的发音差不多时,其计算出来的似然概率也就相近。

参考文献:

[1] 许慧燕,钱盛友.湖南方言声频特性的计算机分析[J].电声技术,2007,31(4):56-58.

[2] Baker W,Eddington D,Nay L.Dialect identification:the effects of region of origin and amount of experience[J].American Speech,2009,84(1):48-71.

[3] 郭春霞.基于 MFCC 的说话人识别系统研究[D].西安:西安电子科技大学,2006.

[4] 赵征鹏,杨鉴.基于高斯混合模型的非母语说话人口音识别[J].计算机工程,2005,31(6):148-150.

[5] 顾明亮,马勇.基于高斯混合模型的汉语方言辨识系统[J].计算机工程与应用,2007,43(3):204-206.

[6] Tsai Wuei-he,Chang Wen-whei.Discrimination training of Gaussian mixture bigram models with application to Chinese dialect identification[J].Speech Communication,2002,36:317-326.

[7] Haranipragada D,Viswwswariah S.Gaussian mixture models with covariances or precisions in shared multiple subspaces[J].IEEE Transactions on Audio Speech and Language Processing,2006,14(4):1255-1266.

(上接 128 页)

表 3 改进多变量贝努里模型的比较实验结果

城市	NP	P	P-IG	P-CHI
样本集 1	93.63%	93.54%	93.54%	93.47%
样本集 2	95.14%	95.44%	95.48%	95.88%
样本集 3	91.15%	93.11%	93.04%	93.92%
样本集 4	93.34%	94.51%	94.36%	94.73%
样本集 5	96.61%	96.56%	96.59%	96.78%
平均值	93.97%	94.63%	94.60%	94.95%

采用了特殊的预处理后,向量的维度平均降低了 68.36%。从表 3 的数据可以看出采用预处理相比没有采用预处理,5 个样本集宏观 *F*1 值的平均值提高了 0.66%,实验证明采用预处理后不仅降低了维度,还提高了分类准确率。在预处理的基础上,采用 IG 和 CHI 两种特征选择算法。实验结果证明,特征选择不仅降低了向量维度,而且不损失分类准确率。P-IG 相比 P,5 个样本集的向量维度平均降低了 48%;P-CHI 相比 P,5 个样本集的向量维度平均降低了 62%。从表 3 的数据看到,P-IG 相比 P,宏观 *F*1 值的平均值几乎相当。P-CHI 相比 *P*,宏观 *F*1 值的平均值提高了 0.32%,说明采用了 CHI 特征选择方法提高了分类的分类准确率。P-CHI 相比 P-IG,宏观 *F*1 值的平均值提高 0.35%。实验结果表明,在犯罪文本分类中,CHI 特征选择在维度的降低和分类准确率两个方面都表现出比 IG 略高的性能。

5 结束语

针对案件文本的特征,提出了特殊的文本预处理方法,比较了两种特征选择方法。并针对案件类别分布不均衡的特点,提出了改进的多变量贝努里模型。实验表明改进的多变量贝努里模型具有较高的分类准确率。

参考文献:

[1] Schneider K M.Techniques for improving the performance of naive bayes for text classification[C]//Proceedings of CICLing,2005:682-693.

[2] Yuan P,Chen Y,Jin H.MSVM-kNN:Combining SVM and k-NN for multi-class text classification[C]//IEEE International Workshop on Semantic Computing and Systems,2008:133-140.

[3] Zhang Haiyi,Li Di.Naive Bayes text classifier[C]//2007 IEEE International Conference on Granular Computing,2007:708-711.

[4] Jie J,Wang G,Qin Y,et al.Crime Data Mining:A general framework and some examples[J].IEEE Computer,2004,37:50-60.

[5] 刘群,张华平,俞鸿魁,等.基于层叠隐马模型的汉语词法分析[J].计算机研究与发展,2004,41(8):1421-1429.

[6] 周茜,赵明生,赵明生.中文文本分类中的特征选择研究[J].中文信息学报,2004,18(3):17-23.

[7] 王维娜,康耀红,伍小芹.文本分类中特征选择方法研究[J].信息技术,2008,32(12):29-31.

[8] McCallum A,Nigam K.A comparison of event model for Naive Bayes text classification[C]//AAAI-98 Workshop on Learning for Text Categorization,1998.