基于概率加权平均的 Mel 子带特征重建算法

罗 宇, 杜利民
(中科院声学所语音交互技术研究中心, 北京 100080)

摘 要: 本文提出基于概率加权平均的 Mel 子带特征数据重建算法. 该算法选择 K 个最优重建结果的概率加权平均作为被加性噪声掩蔽的语音特征分量的估计. 实验结果表明, 基于概率加权平均的语音特征数据重建算法降低了重建误差, 减少了帧间突变现象, 增强了 Mel 子带特征的帧间连续性, 从而显著提高了语音识别系统对加性噪声的鲁棒性能.

关键词: 缺失特征方法; 数据重建; 语音识别
中图分类号: TN912.34 文献标识码: A 文章编号: 0372-2112 (2004) 10-1738-04

Probability-Weighted Average Algorithm for Mel-Frequency Filter-Bank Vector Reconstruction

LUO Yu, DU Li-min
(Center of Speech Interaction Technology Research, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100080, China)

Abstract: In this paper, probability weighted average (PWA) algorithm is proposed to reconstruct Mel-frequency filter-bank vectors. The probability-weighted average of K-best reconstructed / missing-0 components of Mel-frequency filter-bank vectors is taken as the estimation of components masked by additive noise. Experimental results show that PWA algorithm can reduce reconstruction error, increase the continuity between neighbor mel-filterbank vectors and greatly improve automatic speech recognition (ASR) system. s robustness against additive noise.

Key words: missing data method; data reconstruction; robust speech recognition; data reconstruction

1 引言

当语音受噪声干扰时, 语音识别系统性能急剧下降. 缺失特征方法[1~3]是提高语音识别系统噪声鲁棒性的一种方法. 该方法考虑到噪声和语音在时域和频域具有不同能量分布的特性, 把含噪语音谱局部信噪比(Local SNR)低于某个门限的部分标记为/ 缺失0, 即进行缺失分量估计(Missing Components Identification)或者掩蔽估计(Mask Estimation). 经过缺失分量估计后, 可以直接在高信噪比的/ 可靠0部分进行语音识别, 即模型边缘化方法; 也可以先重建出语音特征的缺失分量, 再进行语音识别, 即数据重建方法. 缺失特征方法没有对噪声特性进行假设和限制, 因此能够适用于更多声学环境, 特别是当噪声为不稳定噪声的时候, 缺失特征方法具有更多潜在的优越性.

基于单高斯模型集的缺失特征数据重建算法能够根据/ 可靠0语音特征重建出/ 缺失0语音特征. 重建后的语音特征较好的重现了原始纯净语音特征的形态和分布, 提高了语音识别系统对各类音子的正识率, 增强了语音识别系统对加性噪声的鲁棒性. 另一方面, 基于单高斯模型集的语音特征数据重建算法在重建出完整语音特征的同时, 产生了多种重建误差. 阻碍了语音识别系统性能的进一步提高.

本文提出了基于概率加权平均的语音特征数据重建算法, 通过对比试验研究了该算法对语音识别系统噪声鲁棒性能的改善效果. 同时, 本文研究了候选模型数 K 对基于概率加权平均的语音特征数据重建算法的影响.

论文的第 2 部分分析了基于单高斯模型集的重建算法引起重建误差的原因; 第 3 部分提出了基于概率加权平均的语音特征重建算法; 第 4 部分通过实验, 对比分析了基于概率加权平均的语音特征重建算法对语音识别系统性能的影响; 第 5 部分研究了候选模型个数和语音识别系统性能之间的关系; 论文的第 6 部分给出了最后的结论.

2 基于单高斯模型集的数据重建

本文语音特征重建算法所处理的对象是 Mel 子带特征矢量. 子带特征分析采用在 Mel 频率域均匀分布的 26 个三角滤波器, Mel 子带特征矢量每一维分量都代表了语音信号在对应 Mel 子带内的能量.

收稿日期: 2002-07-18; 修回日期: 2003-11-20
基金项目: 国家 973 重点基础研究发展项目(No. G1998030505)

基于单高斯模型集的重建算法假设纯净语音 Mel 子带特征来自 M 个单高斯模型构成的单高斯模型集. 纯净语音 Mel 子带特征矢量经过 K2mean 聚类和最大似然参数估计得到 M 个单高斯模型参数[13], 并用于 Mel 子带特征重建.

纯净语音特征矢量 S 概率密度函数如公式(1)所示:

$$P_{M_j}(S)=\frac{\exp\left\{-\frac{1}{2}(s-L_j)^{t}H_j^{-1}(s-L_j)\right\}}{(2P)^{n/2}|H_j|^{1/2}} \tag{1}$$

公式(1)中, L_j, H_j 是单高斯模型集中, 第 j 个单高斯模型的均值矢量和协方差矩阵(1[j[M), M 是单高斯模型集合中单高斯模型数目.

经过缺失分量估计后, 语音特征 S 分解为: 语音特征矢量被噪声掩蔽分量构成的/ 缺失矢量0 S^m, 语音特征矢量未被噪声掩蔽分量构成的/ 可靠矢量0 S^o.

基于单高斯模型集的 Mel 子带特征数据重建算法的步骤如下:

首先, 根据/ 可靠矢量0 S^o, 估计语音 Mel 子带特征 S 所属的单高斯模型:

$$j^*=\arg\max_j(P(M_j|S^o))=\arg\max_j\left(\frac{P_{M_j}(S^o)\,@\,P(M_j)}{P(S^o)}\right) \tag{2}$$

其中, $P(S^o)$ 表示出现/ 可靠矢量0 S^o 的概率, 为常数; $P(M_j)$ 是出现第 j 个单高斯模型的先验概率; $P_{M_j}(S^o)$ 模型 M_j 产生/ 可靠矢量0 S^o 的概率, 即第 j 个单高斯模型对/ 可靠矢量0 S^o 的边缘化概率:

$$P_{M_j}(S^o)=\int_{S^m}P_{M_j}(S)\,dS^m=\int_{S^m}P_{M_j}(S^mS^o)\,dS^m \tag{3}$$

由于公式(2)中, $P(S^o)$ 为常数. 所以公式(2)可以简化为:

$$j^*=\arg\max_j(P_{M_j}(S^o)\,@\,P(M_j))=\arg\max_j(P(M_jS^o)) \tag{4}$$

根据第 j* 个单高斯模型参数, 文献[3]提出利用最大后验概率(Maximum A Posteriori)准则来进行缺失特征重建. 缺失特征可以用公式(5)进行估计:

$$S^m=L_{j^*m}+H_{j^*mo}H_{j^*oo}^{-1}(S^o-L_{j^*o}) \tag{5}$$

公式(5)中, L_{j^*m} 表示第 j^* 个单高斯模型中/ 缺失0 分量所对应的均值矢量; L_{j^*o} 表示第 j^* 个单高斯模型中/ 可靠0 分量所对应的均值矢量; H_{j^*oo} 表示第 j^* 个单高斯模型中/ 可靠0 分量所对应的协方差矩阵; H_{j^*mo} 表示第 j^* 个单高斯模型中/ 可靠0 分量和/ 缺失0 分量间的协方差矩阵.

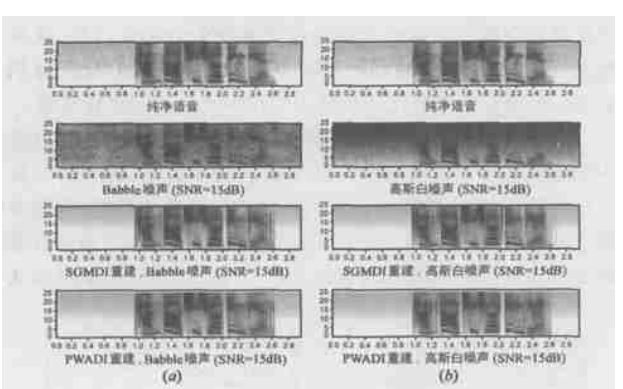

图 1　Mel 特征重建试验结果: (a)受高斯白噪声破坏的语音的实验结果, (b)受 Babble 噪声破坏的语音的实验结果. (图中的汉语语音是: / 谈到汽车定点(tan2 dao4 qi4 che1 ding4 dian3)0)

基于单高斯模型集的 Mel 子带特征重建算法在重建出完整语音 Mel 子带特征矢量的同时, 也引入了多种重建误差[12], 参见图 1. 引起重建误差的主要来源有:

(1)根据/ 可靠矢量0 S^o, 按公式(2)把 S 归类到模型 j^*, 出现归类错误;

(2)Mel 子带特征分布不符合标准单高斯分布.

3　基于概率加权平均的缺失特征数据重建算法

为降低重建误差, 本文提出了基于概率加权平均的缺失特征数据重建算法. 该算法把特征 S 所属高斯模型候选范围从 1 个扩大到 K 个, 并根据候选高斯模型产生/ 可靠矢量0 S^o 的概率, 对 K 个重建特征进行加权平均, 得到/ 缺失矢量0 S^m 的估计.

基于概率加权平均的语音数据重建算法按如下步骤进行:

首先, 根据/ 可靠矢量0 S^o, 估计语音特征 S 在单高斯模型集中产生概率最大的 K 个单高斯模型, 表示为[j_k^*]:

$$\tilde{o}j_k^*8=\arg\max_j k(P_{M_j}(S^o)\,@\,P(M_j))=\arg\max_j k(P(M_jS^o)) \tag{6}$$

其中, $P(M_j)$ 是 S 属于第 j 个单高斯模型的先验概率; $P_{M_j}(S^o)$ 表示模型 M_j 产生语音特征/ 可靠矢量0 S^o 的概率(参见公式(3)).

其次, 根据 K 个单高斯模型[j_k^*]中每个单高斯模型参数, 按照公式(5)重建/ 缺失矢量0

$$S_k^m=L_{j_k^*m}+H_{j_k^*mo}H_{j_k^*oo}^{-1}(S^o-L_{j_k^*o}) \tag{5}$$

重建的 K 个特征矢量表示为[S_k^m]. 计算 K 个重建特征矢量[S_k^m]的概率加权系数:

$$A_k=\frac{P(M_kS^o)}{\sum_{k1[j_k^*]}P(M_kS^o)} \tag{7}$$

K 个概率加权重建系数表示为[A_k]. 最后, 概率加权平均 K 个重建特征矢量[S_k^m], 得到缺失特征矢量估计:

$$S^m=\sum_{k=1}^{K}(A_k\,@\,S_k^m) \tag{8}$$

4　两种重建算法性能的对比及分析

在复杂任务条件(高困惑度非特定人汉语连续语音识别)下, 本部分对比实验分析了基于单高斯模型集的重建算法和基于概率加权平均的语音特征重建算法重建出 Mel 子带特征的差异, 以及两种算法对语音识别系统噪声鲁棒性能的影响.

411　实验条件描述

语音识别系统的训练和测试语音数据均来自 863 语音数据库. 我们使用了其中的 158 人数据进行训练(79 个男声语料, 79 个女声语料), 剩余的 8 人数据用于系统性能测试(4 个

男声语料,4 个女声语料).

语音信号使用 25ms 哈明窗对连续语音进行分帧,并用 0.97 的预加重滤波器提升高频分量,每帧信号重叠 15ms. 子带特征分析采用 0~ 8000Hz 的范围内按照 Mel 刻度均匀分布的 26 通道三角滤波器组. 12 阶 Mel 倒谱,加上对数能量,以及它们的一阶、二阶差分项,构成 39 维特征向量.

HMM 模型结构如下: 停顿模型使用 3 个状态描述; 静音模型使用 5 个状态描述,首尾两个状态没有输出,仅用来连接模型; 每个子音使用 5 个状态描述,首尾两个状态没有输出,仅用来连接模型. 训练得到的系统具有近 9000 个状态,每个状态对应 7 个混合高斯. 解码过程使用文法无关,困惑度为 406 的汉语拼音音节网络.

纯净语音 Mel 子带特征矢量,经过 K2mean 聚类和最大似然参数估计得到 M 个单高斯模型参数[13],并用于 Mel 子带特征重建. 在本文的实验中, M= 256.

含噪语音选用了高斯白噪声混入纯净语音的方式生成,噪声数据来自 NoiseX92 噪声数据库的演示数据. 保存纯净语音和加入的噪声数据用于理想缺失分量估计.

412 Mel 子带特征重建实验

假设纯净语音的 Mel 子带特征矢量为 S,噪声的 Mel 子带特征矢量为 N,理想缺失分量估计按如下公式进行:

$$MSK_i(k)=\begin{cases}1, & \text{if } SNR_i(k)=10\,Log_{10}\left(\dfrac{S_i(k)}{N_i(k)}\right) > D\\ 0, & \text{if } SNR_i(k)=10\,Log_{10}\left(\dfrac{S_i(k)}{N_i(k)}\right) \le D\end{cases} \quad (9)$$

公式(9)中, $MSK_i(k)=1$ 表示第 i 帧语音第 k 个 Mel 子带/ 可靠0; $MSK_i(k)=0$ 表示第 i 帧语音第 k 个 Mel 子带受噪声掩蔽/ 缺失0; $S_i(k)$ 是纯净语音第 i 帧 Mel 子带特征的第 k 个分量,对应第 k 个 Mel 三角子带内的纯净语音能量; $N_i(k)$ 是噪声第 i 帧 Mel 子带特征的第 k 个分量,对应第 k 个 Mel 三角子带内的噪声能量; R 是判断特征是否可靠的门限值,根据人耳掩蔽效应,选择 R 为- 5dB~ 5dB.

理想缺失分量估计的条件在实际环境中很难得到满足,因为很难得到孤立的噪声信号和语音信号. 本文利用理想缺失分量估计来评价缺失特征重建算法的效果.

图 1 给出加入高斯白噪声前后语音的 Mel 子带特征图(x 轴为时间轴, y 轴为 Mel 子带轴)以及理想缺失分量估计后,利用基于单高斯模型集的重建算法(表示为 SGMDI)和基于概率加权平均的重建算法(表示为 PWADI)分别重建后,得到的 Mel 子带特征图.

从 Mel 子带特征图可以看出,加性噪声改变了纯净语音特征矢量的形态和分布,造成语音识别系统识别率大大下降. 缺失特征重建算法能够重建出受噪声破坏的 Mel 子带特征,重建后的 Mel 子带特征较好的重现了原始纯净语音段 Mel 子带特征的形态和分布.

对比基于单高斯模型集的缺失特征数据重建算法,基于概率加权平均的缺失特征数据重建算法重建得到的 Mel 子带特征图帧间连续性较好,减少了缺失特征数据重建过程中的帧间突变现象(参见图 1).

413 语音识别实验

表 1 给出了经过理想缺失分量估计和数据重建(K= 12)处理后语音识别系统的音节识别性能. 从表 1 可以看出,经过数据重建后,由于重建后的美子带特征图较好的重现了原始纯净语音段美子带特征的形态和分布,因此系统识别率有了较大的提高.

表 1 音节识别性能比较(SGMDI 算法/ PWADI 算法)

含噪语音		音节正确率(%)			音节准确率(%)		
噪声类型	SNR(dB)	NOISY	SGMDI	PWADI	NOISY	SGMDI	PWADI
Babble 噪声	0	3.09	23.78	25.68	- 3.07	13.64	15.70
	5	9.72	41.75	44.37	- 5.81	32.03	35.08
	10	27.53	58.41	59.91	8.15	51.23	53.56
	15	48.62	68.20	69.12	29.74	62.36	63.72
	20	62.11	73.88	74.11	45.97	68.75	69.10
	25	70.48	76.13	76.41	58.91	71.49	72.09
高斯白噪声	0	2.80	15.80	18.06	2.49	3.56	7.12
	5	7.06	32.19	34.99	2.34	20.24	25.27
	10	17.19	48.07	51.23	1.31	39.49	44.21
	15	32.31	57.84	60.96	12.09	51.02	55.32
	20	48.28	66.35	68.26	28.00	60.17	63.08
	25	61.19	72.33	73.26	56.14	67.25	68.86

基于概率加权平均的语音数据重建算法通过扩大候选模型,并对候选模型重建得到的特征矢量按照产生概率加权平均,减轻归类错误和模型分布不符合高斯分布的影响. 在不同信噪比不同噪声类型环境下,性能均优于基于单高斯模型集的数据重建算法的性能.

5 模型候选数 K 对基于概率加权平均的语音数据重建算法的影响

为了考察候选模型数 K 对语音识别系统性能的影响,选择受高斯白噪声破坏的语音经过理想缺失分量估计后,在其他条件完全一致的情况下,调整候选模型数 K 的取值,分别对 800 个语音文件进行基于概率加权平均的数据重建实验.

图 2 给出了实验结果. 随着候选模型数 K 的增加,语音识别系统的音节正确率和音节准确率逐步提高,说明增加候选模型数 K 能够提高语音识别系统的噪声鲁棒性能. 但随着候选模型数 K 的增加,语音识别系统的音节正确率和音节准确率逐步的提高幅度逐渐变缓. 另一方面,随着候选模型数 K

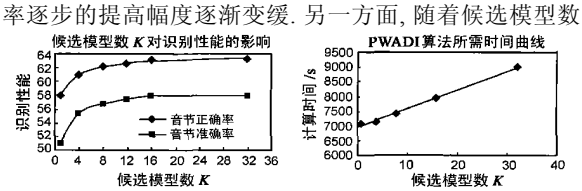

图 2 候选模型数 K 对系统识别性能的影响
(单高斯模型数 M= 256)

的增加,语音特征重建运算量也相应增加,不利于该算法的实时实现(参见图 2).

6　结论

基于概率加权平均的缺失特征重建算法增加候选模型,并对多个候选模型重建结果基于概率加权平均,降低了模型误判和特征矢量分布不符合高斯分布的影响,减少了缺失特征数据重建过程中的帧间突变现象,增强了 Mel 子带特征的帧间连续性,从而提高了语音识别系统对加性噪声的鲁棒性能.

参考文献:

[1] A Vizinho, P Green, M Cooke, L Josifovski. Missing data theory, spectral subtraction and signal2to2noise estimation for robust ASR: An integrated study[A]. Eurospeech. 99[C]. Budapest, 1999.

[2] Martin Cooke, Phil Green, Ljubomir Josifovski and Ascension Vizinho. Robust ASR with unreliable data and minimal assumptions[A]. Robust 99[C]. Tamper, Finland.

[3] Morris A C, Cooke M, Green P. Some solutions to the missing feature problem in data classification, with application to noise robust ASR [A]. Proc. ICASSP. 98[C]. 1998. 737- 740.

[4] B Raj, M L Seltzer, R M Stern. Robust speech recognition: the case for restoring missing features[A]. Proceedings of the Workshop on Consistent and Reliable Acoustic Cues for Sound Analysis (CRAC) 2001 [C]. September, 2001, Aalborg, Denmark.

[5] Ljubomir Josifovski, Martin Cooke, Phil Green, Ascension Vizinho. State based imputation of missing data for robust speech recognition and speech enhancement[A]. in Eurospeech[C]. 1999. 6. 2833- 2836.

[6] Jon Barker, Ljubomir Josifovski, Martin Cooke and Phil Green. Soft decisions in missing data techniques for robust automatic speech recognition[A]. ICSLP22000, Beijing[C]. 2000. 373- 376.

[7] B Raj, M L Seltzer, R M Stern. Reconstruction of damaged spectrographic features for robust speech recognition[A]. Proceedings of the International Conference on Spoken Language Processing[C]. October, 2000, Beijing, China.

[8] M L Seltzer, B Raj, R M Stern. Classifier2based mask estimation for missing feature methods of robust speech recognition[A]. Proceedings of the International Conference on Spoken Language Processing[C]. October, 2000, Beijing, China.

[9] Philippe Renevey, Rolf Vetter, Jens Kraus. Robust speech recognition using missing feature theory and vector quantization[A]. Eurospeech 2001[C]. Scandinavia, 2001. 1107.

[10] Lippmann R P, Carlson B A. Using missing feature theory to actively select features for robust speech recognition with interruptions, filtering and noise[A]. Proc Eurospeech. 97 [C]. Rhodes, Greece, September 1997. KN372 40.

[11] Steve Young, Dan Kershaw, Julian Odell, Dave Ollason, Valtcho Valtchev, Phil Woodland, The HTK Book (for HTK Version 3. 1) [M].

[12] B Raj. Reconstruction of Incomplete Spectrograms for Robust Speech Recognition, Ph. D Dissertation[D]. ECE Department, CMU, April, 2000.

[13] 边肇祺, 张学工 等编著. 模式识别[M]. 清华大学出版社, 2001 年 1 月第 2 版.

作者简介:

罗　宇　(见本期第 1657 页)

杜利民　(见本期第 1657 页)