第 26 卷第 5 期
2004 年 5 月

电 子 与 信 息 学 报
Journal of Electronics & Information Technology

Vol.26No.5
May 2004

基于隐马尔可夫模型局部最优状态路径的数据重建算法 [1]

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：　该文提出了基于隐马尔可夫模型局部最优状态路径的数据重建 (LOPDI) 算法. 该算法假设语音特征矢量是一个 L 状态隐马尔可夫模型的输出序列，基于局部最优状态路径估计产生语音特征矢量的次最优状态序列，并按最大后验概率准则 (MAP) 重建出“缺失矢量”. 实验表明，LOPDI 算法能够显著提高语音识别系统对加性噪声的鲁棒性.

关键词：　LOPDI, 缺失特征方法，数据重建，隐马尔可夫模型

中图分类号：　TN912.3　　**文献标识码**：　A　　**文章编号**：　1009-5896(2004)05-0722-05

HMM Local Optimal State Path-Based Data Imputation

Luo Yu　　Du Li-min

(*Center of Speech Interaction Technology Research,*

Institute of Acoustics, Chinese Academy of Sciences, Beijing 100080, *China*)

Abstract　This paper presents a HMM Local Optimal state Path-based Data Imputation (LOPDI) algorithm. Speech feature vector sequences are presumed to be the outputs of an L state HMM. The HMM state sequence is estimated by local optimal path procedure. Then, "missing" data is recovered by MAP procedure. Experimental result shows that LOPDI algorithm can greatly increase automatic speech recognition system's robustness against additive noise.

Key words　LOPDI, Missing feature method, Data imputation, HMM

1 引言

缺失特征方法 [1−3] 认为噪声和语音在时间 - 频率域上不同区域具有不同局部信噪比，并进行缺失分量估计 (Missing component identification) 或者掩蔽估计 (Mask estimation)，即把局部信噪比较低的区域标记为“缺失”，而局部信噪比较高的区域标记为“可靠”。经过缺失分量估计后，可以根据“可靠矢量”进行语音识别，即模型边缘化方法；也可以重建“缺失矢量”，得到完整矢量后进行语音识别，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，当噪声为不稳定信号时，该方法具有潜在的优越性。

导致基于单高斯模型集的数据重建算法 (SGMDI) 算法 [4] 出现重建误差的主要原因是归类错误 [5]。为了减小归类错误的影响，本文引入隐马尔可夫模型来描述相邻语音特征矢量统计信息，提出了基于隐马尔可夫模型局部最优状态路径的数据重建 (LOPDI) 算法，考察了该算法提高语音识别系统噪声鲁棒性的作用。

论文的第 2 节介绍理想缺失分量估计方法；第 3 节提出了 LOPDI 算法；第 4 节通过实验对比分析了 SGMDI 算法和 LOPDI 算法的性能；第 5 节给出了最后的结论。

2 理想缺失分量估计方法

本文数据重建算法所处理的对象是美子带 (Mel-filter-bank) 特征矢量。缺失分量估计和数

[1] 2003-02-27 收到，2003-06-10 改回
国家 973 重点基础研究发展项目资助课题 (G1998030505)

据重建在美子带特征矢量空间内进行．语音信号经过在美频率 (Mel-frequency) 域均匀分布的三角滤波器组进行子带特征分析，得到美子带特征矢量序列。假设纯净语音美子带特征矢量为 $\boldsymbol{S}$，噪声美子带特征矢量为 $\boldsymbol{N}$，理想缺失分量估计按如下公式进行：

$$\mathrm{MSK}_i(k)=\begin{cases}1, & \hat{\mathrm{SNR}}_i(k)=10\lg\left(\dfrac{S_i(k)}{N_i(k)}\right)>\delta\\ 0, & \hat{\mathrm{SNR}}_i(k)=10\lg\left(\dfrac{S_i(k)}{N_i(k)}\right)\le\delta\end{cases}\tag{1}$$

其中 $S_i(k)$, $N_i(k)$ 分别是纯净语音和噪声第 i 帧美子带特征的第 k 个分量 (对应第 k 个美三角子带内的纯净语音能量和噪声能量)； δ 是判断该子带是否可靠的门限，根据人耳掩蔽效应，选择 δ 范围为 $-5\mathrm{dB}\sim5\mathrm{dB}$．掩蔽矢量 MSK 描述了噪声对语音的掩蔽情况： $\mathrm{MSK}_i(k)=1$ 表示第 i 帧语音第 k 个美子带信噪比较高，是“可靠”子带； $\mathrm{MSK}_i(k)=0$ 表示第 i 帧语音第 k 个美子带信噪比较低，是“缺失”子带。经过缺失分量估计，语音特征 $\boldsymbol{S}$ 分为两个矢量：由不可靠分量构成的“缺失矢量” $\boldsymbol{S}^m$ 及由可靠分量构成的“可靠矢量” $\boldsymbol{S}^o$．

理想缺失分量估计的条件在实际环境中很难得到满足，因为很难准确分离噪声信号和语音信号。但是，理想缺失分量估计可以用于评价数据重建算法的性能。

3 基于隐马尔可夫模型局部最优状态路径的数据重建 (LOPDI)

3.1 隐马尔可夫模型 [6]

假设一个离散时域有限状态自动机，在每一离散时刻 t，自动机所处的状态用 x_t 表示，有 $x_t\in[Q_j]$，其中 $[Q_j]=[Q_1,\cdots,Q_L]$，表示 L 个可能出现的状态。假设自动机开始时刻 $t=1$，则在以后每一个时刻 $t>1$，自动机所处的状态以概率方式取决于初始状态概率矢量 $\boldsymbol{a}$ 和状态转移概率矩阵 $\boldsymbol{A}$．

初始状态概率矢量 $\boldsymbol{a}$ 是一个 L 维矢量， $\boldsymbol{a}=[a_1,\cdots,a_L]$，初始状态概率 a_i 表示在开始时刻，自动机处于状态 Q_i 的概率：

$$a_i=P(x_1=Q_i),\quad 1\le i\le L\tag{2}$$

状态转移概率矩阵 $\boldsymbol{A}$ 是一个 $(L\times L)$ 维方阵，状态转移概率 A_{ij} 表示在相邻两个时刻系统状态从 Q_i 转移到 Q_j 的概率：

$$A_{ij}=P(x_t=Q_j/x_{t-1}=Q_i),\quad t>1,\ 1\le i,\ j\le L\tag{3}$$

显然有

$$\sum_{j=1}^{L}A_{ij}=1,\quad \forall i,\ 1\le i,\ j\le L\tag{4}$$

对于任何 $t>1$ 时刻，自动机所处状态 x_t 只取决于前一时刻所处的状态 x_{t-1}．从时刻 1 到时刻 T，状态序列 $[x_1,x_2,\cdots,x_T]$ 构成了一条一阶马尔可夫链。

在任意时刻，当系统处于状态 Q_i 时，观测到美子带特征 $\boldsymbol{S}$ 的概率表示为

$$b_i(\boldsymbol{S})=P_{Q_i}(\boldsymbol{S})=P(\boldsymbol{S}/x=Q_i),\quad 1\le i\le L\tag{5}$$

L 个状态的概率分布构成一个 L 维矢量 $\boldsymbol{B}$，表示为

$$\boldsymbol{B}=[b_1(\boldsymbol{S}),\cdots,b_i(\boldsymbol{S}),\cdots,b_L(\boldsymbol{S})]\tag{6}$$

3.2 基于局部最优状态路径的数据重建

基于隐马尔可夫模型数据重建的实质是根据隐马尔可夫模型参数 $[a, \boldsymbol{A}, \boldsymbol{B}]$ 和“可靠矢量”序列 $[\boldsymbol{S}_1^o, \boldsymbol{S}_2^o, \cdots, \boldsymbol{S}_T^o]$，估计“缺失矢量”序列 $[\boldsymbol{S}_1^m, \boldsymbol{S}_2^m, \cdots, \boldsymbol{S}_T^m]$。本文提出基于隐马尔可夫模型的 LOPDI 算法：该算法根据局部最优状态路径估计产生语音特征矢量的次最优状态序列，并基于最大后验概率准则，重建“缺失矢量”。

LOPDI 算法按如下步骤进行：

(1) 初始化

在初始时刻 1，估计系统所处的初始状态：

$$x_1 = \arg \max_{x_i \in [Q_j]} [a_j b_j(\boldsymbol{S}_1^o] \tag{7}$$

其中 a_j 表示初始状态为状态 j 的概率，$b_j(\boldsymbol{S}_1^o)$ 表示系统处于状态 j 的情况下，观测到“可靠矢量” $\boldsymbol{S}_1^o$ 的概率，即 $b_j(\boldsymbol{S}_1)$ 对 $\boldsymbol{S}_1^o$ 的边缘化概率：

$$b_j(\boldsymbol{S}_1^o) = P_{Q_j}(\boldsymbol{S}_1^o) = \int P_{Q_j}(\boldsymbol{S}_1) \mathrm{d}\boldsymbol{S}^m = \int P_{Q_j}(\boldsymbol{S}_1^o \boldsymbol{S}^m) \mathrm{d}\boldsymbol{S}^m \tag{8}$$

$\boldsymbol{S}_1$ 所属状态 x_1 确定后，基于最大概率准则重建“缺失矢量”。即估计“缺失矢量” $\boldsymbol{S}_1^m$，使产生语音特征矢量 $(\boldsymbol{S} = [\boldsymbol{S}_1^o \boldsymbol{S}_1^m])$ 的概率 $b_{x_1}(\boldsymbol{S}_1^o \boldsymbol{S}_1^m)$ 最大：

$$\hat{\boldsymbol{S}}_1^m = \arg \max_{\boldsymbol{S}^m} [b_{x_1}(\boldsymbol{S}_1^o \boldsymbol{S}_1^m)] \tag{9}$$

假设 $b_i(\boldsymbol{S})$ 符合高斯分布，即

$$b_i(\boldsymbol{S}) = P_{Q_i}(\boldsymbol{S}) = \exp[-(1/2)(\boldsymbol{S} - \boldsymbol{\mu}_i)^t \boldsymbol{\Theta}_i^{-1} (\boldsymbol{S} - \boldsymbol{\mu}_i)] / [(2\pi)^{n/2} |\boldsymbol{\Theta}_i|^{1/2}] \tag{10}$$

其中 n 是语音特征矢量维数，$\boldsymbol{\mu}_i$, $\boldsymbol{\Theta}_i$ 是隐马尔可夫模型中第 i 个状态的均值矢量和协方差矩阵 $(1 \le i \le L)$。求解式 (9)，得到 [5]：

$$\hat{\boldsymbol{S}}_1^m = \boldsymbol{\mu}_{x_1 m} + \boldsymbol{\Theta}_{x_1 mo} \boldsymbol{\Theta}_{x_1 oo}^{-1} (\boldsymbol{S}^o - \boldsymbol{\mu}_{x_1 o}) \tag{11}$$

其中 x_1 表示系统的初始状态；$\boldsymbol{\mu}_{x_1 o}$ 表示状态 x_1 下，“可靠矢量”对应的均值矢量；$\boldsymbol{\mu}_{x_1 m}$ 表示状态 x_1 下，“缺失矢量”对应的均值矢量；$\boldsymbol{\Theta}_{x_1 oo}$ 表示状态 x_1 下，“可靠矢量”的协方差矩阵；$\boldsymbol{\Theta}_{x_1 mo}$ 表示状态 x_1 下，“可靠矢量”和“缺失矢量”间的协方差矩阵。

(2) 局部最优状态路径估计

对每个时刻 $t > 1$，状态 x_t 由前一状态 $x_{t-1} = Q_j$，转移概率 A_{ij} 和在时刻 t 观测到的“可靠矢量” $\boldsymbol{S}_t^o$ 决定：

$$x_t = \arg \max_{x_i \in [Q_j]} [A_{ij} * b_j(\boldsymbol{S}_t^o)], \qquad 2 \le t \le T \tag{12}$$

(3) 缺失分量重建

$\boldsymbol{S}_t$ 所属状态 x_t 确定后，基于最大概率准则重建“缺失矢量” $\boldsymbol{S}_t^m$ (参见式 (9), (10), (11))：

$$\hat{\boldsymbol{S}}_t^m = \boldsymbol{\mu}_{x_1 m} + \boldsymbol{\Theta}_{x_1 mo} \boldsymbol{\Theta}_{x_1 oo}^{-1} (\boldsymbol{S}^o - \boldsymbol{\mu}_{x_1 o}) \tag{13}$$

(4) 重复步骤 (2)~(3)，直到语音输入结束.

下面，本文将在大词汇表非特定人汉语连续语音识别任务下，通过试验对比分析 LOPDI 算法的重建效果，以及对提高语音识别系统噪声鲁棒性的作用。

4 试验分析

4.1 实验条件

大词汇表非特定人汉语连续语音识别系统的训练和测试数据来自 863 语音数据库。噪声数据来自 NoiseX-92 噪声数据库。试验选用 2 种噪声 (高斯白噪声，Babble 噪声)，按照 SNR=15dB 加入纯净语音，同时，保存纯净语音和噪声用于理想缺失分量估计。

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，相邻帧重叠 15ms 。语音特征矢量选择 MFCC-E-D-A 。子带分析采用在 0~8000Hz 范围内按照美 (Mel) 刻度均匀分布的 26 通道三角滤波器组。语音识别系统隐马尔可夫模型的结构为：停顿模型使用 3 个状态；静音和子音模型使用 5 个状态，首尾两个状态没有输出，仅用来连接模型。解码使用文法无关，困惑度为 406 的汉语拼音音节网络。

受高斯白噪声破坏的语音首先转换为美子带特征，经过理想缺失分量估计、数据重建后，转换为 MFCC(Mel Frequency Cepstral Coefficient) 特征，并作为识别器的输入。试验中，理想缺失分量估计选择 $\delta = -5\text{dB}$ ； SGMDI 算法选择单高斯模型数 $N = 256$ ； LOPDI 算法选择隐马尔可夫模型状态 $L = 256$ 。

4.2 实验结果及分析

试验结果表明，加性噪声破坏了纯净语音特征矢量的形态和分布。数据重建算法能够有效地重建出受加性噪声破坏的美子带特征，重建后的美子带特征较好地重现了原始纯净语音美子带特征的形态和分布。

LOPDI 算法利用隐马尔可夫模型描述相邻美子带特征之间的统计信息，因此重建后的美子带特征具有较好的连续性，减少了由于相邻美子代特征之间的不连续导致的“奇异”特征。另一方面， LOPDI 算法可能陷入局部最优状态，无法跳转到正确状态，表现为语音结束后，连续重建出相同的非静音子带特征。

在语音识别实验中，定义识别正确率 (Correct) 和准确率 (Accuracy)[7] 为

$$\text{正确率} = \frac{N - D - \text{Sub}}{N} \times 100\%$$

$$\text{准正确率} = \frac{N - D - \text{Sub} - I}{N} \times 100\% \tag{14}$$

其中 N 为识别单元 (单词 (word) 、音子 (phoneme) 、音节 (syllable) 等) 出现的总次数， D 为删除错误次数， Sub 为替换错误次数， I 为插入错误次数。

实验结果表明，加性噪声的存在将导致语音识别系统对各类音子的识别正确率发生明显下降：能量较低，持续时间较短的清辅音容易受到噪声的破坏，音子识别正确率大幅下降；能量较高，持续时间较长的浊辅音和元音抵抗噪声能力较强，音子识别正确率下降幅度较小。经过理想缺失分量估计、数据重建后，各类音子的平均识别正确率得到明显提高。相对于 SGMDI 算法， LOPDI 算法能够进一步提高各类音子的平均识别正确率 (参见表 1) 。

表 1　音子平均识别正确率分类统计表

音子分类	受 Babble 噪声破坏语音			受高斯白噪声破坏语音		
	$\delta\%c$	$\Delta\%c_1$	$\Delta\%c_2$	$\delta\%c$	$\Delta\%c_1$	$\Delta\%c_2$
清辅音 (%)	−28.22	20.86	21.46	−49.06	27.46	31.89
浊辅音 (%)	−10.87	7.40	8.10	−32.93	23.23	23.77
元音 (%)	−15.18	10.22	10.69	−25.59	15.77	19.31

注： $\delta\%c$ 表示噪声引起该类音子平均识别正确率对纯净语音的改变量，
如“ −28.22 ”表示噪声引起该类音子平均识别正确率比纯净语音下降了 28.22% ；
$\Delta\%c_1$ 表示 SGMDI 重建后，该类音子平均识别正确率对含噪语音的改变量，
如“ 20.86 ”表示经过 SGMDI 重建后，该类音子平均识别正确率比含噪语音上升了 20.86% ；
$\Delta\%c_2$ 表示 LOPDI 重建后，该类音子平均识别正确率对含噪语音的改变量，
如“ 21.46 ”表示经过 LOPDI 重建后，该类音子平均识别正确率比含噪语音上升了 21.46% ；

音节由音子结合而成，是语音的最基本组成单位。音节层次的识别性能在能够有效地衡量大词汇表非特定人连续语音识别系统的性能。实验结果表明，加性噪声破坏了语音特征矢量的形态和分布，造成语音识别系统性能大幅下降。数据重建减少了语音特征矢量畸变，语音识别系统性能有了显著提高。而且，相对于 SGMDI 算法，采用 LOPDI 算法使语音识别系统的性能有了进一步的提高 (参见表 2) 。

表 2 含噪语音美子带特征重建，语音识别实验结果比较 (高斯 /Babble 噪声， SNR=15dB)

噪声类型	识别性能	纯净语音	含噪语音	SGMDI(N=256)	LOPDI(L = 256)
高斯白噪声	音节正确率 (%)	78.62	32.31	57.84	62.08
	音节准确率 (%)	74.71	12.09	51.02	55.89
Babble 噪声	音节正确率 (%)	78.62	48.62	68.20	68.85
	音节准确率 (%)	74.71	29.74	62.36	62.63

5 结论

LOPDI 算法基于隐马尔可夫模型局部最优状态路径估计产生语音特征矢量的次最优状态序列，并基于最大概率准则完成“缺失矢量”重建。实验表明，隐马尔可夫模型能够更好地描述相邻语音特征间的统计信息，从而减少了相邻重建语音特征间的不连续现象，提高了语音识别系统对加性噪声的鲁棒性。

参 考 文 献

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

[2] Vizinho A, Green P, Cooke M, Josifovski L. Missing data theory, spectral subtraction and signal-to-noise estimation for robust ASR: An integrated study. Eurospeech'99, Budapest, 1999, vol.5: 2407–2410.

[3] Cooke M, Green P, Josifovski L, Vizinho A. Robust automatic speech recognition with missing and unreliable acoustic data. *Speech Communication*, 2001, 34(3): 267–285.

[4] Raj B, Seltzer M, Stern R. Reconstruction of damaged spectrographic features for robust speech recognition. In Proceedings ICSLP'00, Beijing, China, October 2000, vol.1: 375–360.

[5] Raj B. Reconstruction of incomplete spectrograms for robust speech recognition. [Ph.D Thesis], ECE Department, Carnegie Mellon University, April, 2000.

[6] Rabiner L R, Juang B H. Fundamentals of Speech Recognition. Prentice-Hall Press, 1993, Ch.6: 321–389.

[7] Steve Young, Dan Kershaw, Julian Odell, Dave Ollason, Valtcho Valtchev, Phil Woodland. The HTK Book (for HTK Version 3.0), Microsoft Corporation, 2000.

罗 宇： 男， 1974 年生，博士生，研究方向为语音交互技术、鲁棒语音识别技术等.

杜利民： 男， 1957 年生，研究员，博士生导师，从事语音信号与信息处理技术的研究，主持研究的主要项目包括连续语音鲁棒识别、连续语音关键语检测、语音交互助理、自然口语识别与对话、语音同声翻译、噪声环境下语音增强和语音提取、低速率语音压缩.