一种改进的线性区分分析方法及其在汉语数码语音识别上的应用

史媛媛, 刘 加, 刘润生
(清华大学电子工程系, 北京 100086)

摘 要: 尽管汉语数码语音识别只涉及十个数字, 但由于不同数字的发音存在相同或相似的声母或韵母, 造成汉语数码语音之间的混淆性很大. 采用通常的隐含马尔科夫模型(HMM) 作为汉语数码语音识别模型难以得到很高的识别率. 为了解决汉语数码之间的混淆问题, 提高汉语数码语音识别性能, 本文在隐含马尔科夫模型的状态层次上采用线性区分分析方法, 将不同状态之间容易混淆的特征样本构成混淆模式类, 针对混淆模式类进行线性区分分析. 通过线性区分变换, 在变换特征空间中仅保留那些能够有效区分该混淆类别的特征参数. 这种基于状态的线性区分分析有效地提高了模型对混淆数码的区分能力. 实验表明即使采用状态数很少的粗糙识别模型, 也能很大幅度提高模型的识别性能; 经过线性区分变换优化后的汉语数码识别模型, 孤立汉语数码语音识别率可以达到 99. 32%.

关键词: 线性区分分析 (LDA); 汉语数码语音识别; 区分变换
中图分类号: TN912 文献标识码: A 文章编号: 0372-2112 (2002) 07-0959-05

An Improved Linear Discriminant Analysis for Mandarin Digit Speech Recognition

SHI Yuan-yuan, LIU Jia, LIU Run-sheng
(Department of Electronic Engineering, Tsinghua University, Beijing 100084, China)

Abstract: It is found that the phonetic similarities in the Mandarin digits are the main reasons for the difficulty of Mandarin digit recognition. In this paper, an improved linear discriminant analysis (LDA) based on the states of hidden Markov models (HMM) is presented. The recognition model discriminability is greatly improved by gathering the confusion data to the given states and then using the state-specific discriminative transformation. The experiments show that it increases the recognition rate greatly even if the simple models with insufficient states are used. The recognition accuracy of isolated Mandarin digits is over 99. 32% after using optimal linear discriminative transformation.

Key words: LDA; Mandarin digit speech recognition; discriminative transformation

1 引言

“数码” 是人们日常交互与信息查询中最重要的输入信息之一. 数码语音识别无论在大词表语音识别系统, 还是在小词表语音识别系统中都占有重要地位, 具有重要的应用价值. 由于汉语数码语音发音短, 并且存在易混语音集合, 因此汉语数码语音的识别比英文数码语音要困难. 如何提高汉语数码音识别性能一直是语音识别中研究热点之一. 进行语音识别特征参数选择和变换, 从中挑选出区分能力较强的特征分量, 去掉冗余的特征分量, 是提高识别模型的区分性能的重要方法之一. 线性区分分析(LDA) 方法是其中一种提高参数区分能力的有效方法, 也是模式识别理论提取鉴别特性的重要理论之一.

本文将线性区分分析用于汉语数码语音识别系统, 改进汉语数码语音识模型的鉴别能力. 线性区分分析采用线性变换矩阵将原始的 n 维特征矢量线性变换成 m 维 ($m < n$) 特征矢量. 通过对特征空间的线性区分变换, 可以消除特征分量的线性相关性; 从变换后的特征参数中提取出最有效的识别特征分量, 去掉没有鉴别能力的特征分量, 提高特征的鉴别性能; 变换空间中的类间散度对类内散度的比率增加, 提高了类别的可分离性[1]. 此外, 特征矢量的维数降低, 有利于减少对训练语音数据的要求, 使识别模型得到充分的训练, 有利于提高识别系统的稳健性.

Hunt 首先将线性区分分析用于语音识别[2, 3], 随后相当多的语音识别系统使用了线性区分分析, 并且不同系统的识

收稿日期: 2000-11-21; 修回日期: 2001-09-28
基金项目: 国家自然科学基金(No. 69975007); 国家 863 项目(No. 863-306ZD13-04-6)

别性能都有了一致性的提高[4~6],尤其是对高混淆度的语音识别任务,如字母表 E 族字母的识别,线性区分分析有效地提高了识别率[7].

在汉语语音识别任务中,汉语数码的识别同样也是一个易混淆识别问题.由于所有汉语数码的发音都是单音节,并且包括了几组易混淆数码,如"6/liu/,9/jiu/","1/yi/,7/qi/","2/er/,8/ba/".这使得目前汉语数码的识别率仍然相对较低,无法满足实际应用的需要.本文将线性区分分析应用于汉语数码语音识别,为了最有效地提高汉语数码模型对易混淆数码的区分能力,在 HMM 模型(声学模型)的状态层次进行线性区分变换.首先在识别模型的各个状态上,采集易与其混淆的其它所有状态的特征样本集合,然后进行线性区分分析.通过特征线性变换,有效地提高了变换特征在特定状态对易混数据集合的区分性能,从而提高识别模型的识别性能.

2 汉语数码识别基线系统

基线汉语数码识别系统语音分析帧长为 24 毫秒,帧移 12 毫秒.提取特征矢量包括 12 阶线性预测倒谱系数,12 个一阶差分系数,以及归一化对数能量特征及其一阶差分,一共 26 维特征参数.系统声学模型包括 15 个基于音素的连续隐含马尔科夫模型以及 1 个静音模型(BG).15 个音素模型包括 6 个声母模型和 9 个韵母模型,分别对应汉语数码发音的声母和韵母部分.每个数码的声学模型由相应的声母模型和韵母模型拼接而成.在训练语音库数据有限的情况下,为了减少识别模型参数的数量,构建模型时未考虑音节内的上下文发音的相关性.静音模型只有一个状态,其它音素模型的状态数根据发音的平均时长不同,分别取 2 个状态到 8 个状态不等,状态的特征样本输出函数采用满协方差阵高斯分布函数.模型训练采用改进 Baum-Welch 算法,训练中没有考虑状态之间的跳转概率.图 1 给出了所有数字的模型连接情况.其中数字"1"包括[yi]和[yao]两种发音,发音[yi]标记为"1",发音[yao]标记为"A".声母和韵母的汉语拼音和国际音标标注列在附录中.

汉语数码语音库包括 160 人发音,其中男女性各 80 人.每人对十个汉语数码的 11 种发音各读一遍.识别率测试时用 120 人做训练集,其余 40 人做测试集,在训练语音集合与测试语音集合无交迭情况下测试 4 次,4 次测试结果的平均值作为最终的识别率结果.

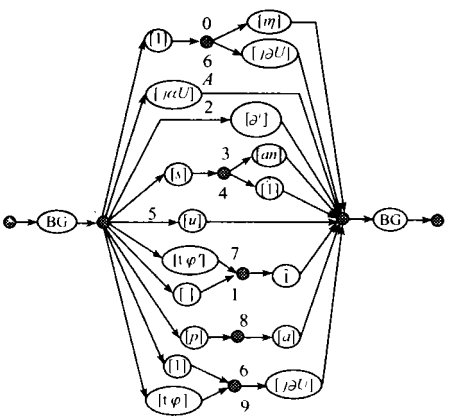

图 1 汉语数码的声母模型和韵母模型

3 线性区分分析

3.1 针对汉语数码状态混淆集合的线性区分分析

不同汉语数码语音之间存在相同或相近的声母和韵母是汉语数码识别高混淆度的主要原因.表 1 总结了数码的发音相似性.采用线性区分分析可以有效提高识别模型对混淆集合的区分能力.线性区分分析可以在音节,音素,隐含马尔科夫模型的状态,甚至状态的高斯混合类上来进行.在不同层次上定义区分类别,对线性区分分析的效果有不同的影响.由于汉语数码的发音相似性主要表现为存在相近的音节,以及相似的音素,在基线系统的识别模型结构中,如何充分提高连续隐含马尔科夫模型状态的高斯分布特征参数的鉴别性能是至关重要的,因此本文将线性区分分析定义在 HMM 的状态层次上.同时,借鉴了文[4,7,9,10]中提出的混淆数据分析方法,用高斯状态分布的协方差矩阵代替线性区分分析中使用的平均类内散度矩阵,用该状态对应混淆类别的高斯分布协方差矩阵代替平均类间散度矩阵.实验表明该方法可以有效地提高线性区分分析特定状态对混淆类别鉴别的性能.

表 1 混淆数码的声母和韵母的发音相似性

混淆数码	声母部分的相似性	韵母部分的相似性
0-6	边音[l]相同	滑音[j]与对应的前高位元音[i]十分相似
A-6	——	韵母中包涵了相同的发音[j]和[U]
9-6	——	韵母[jU]相同
3-4	摩擦音[s]相同	——
7-4	塞擦音[tɕʻ]和摩擦音[s]有语音相似性	[i]和[î]是相似的前高位元音
7-1	——	韵母[i]相同
2-8	塞音[p]的发音时长很短,不易识别	[ɔʳ]易被发成与[a]相似的音

3.2 状态混淆集合

为了在模型的状态层次上使用线性区分分析,首先需要得到待分析的状态类别及其混淆类别的特征样本集合.目前主要的混淆数据采集包括两种方式:

一种混淆数据采集方式根据各帧特征矢量在特征空间的"距离"大小来决定是否存在混淆.这类方式一般考察各帧特征矢量与错误状态中心的"距离"是否小于其对应的正确状态中心的"距离",如果某帧语音特征与其它状态中心的"距离"和与正确状态中心的"距离"的差值小于一定阈值时,认为该帧属于其它状态的混淆集合."距离"可以是语音识别中普遍使用的欧式距离,马氏距离,或是其它统计距离,或者是状态对该特征帧的输出概率分布函数值.

另一种混淆数据采集方式基于 Viterbi 搜索的结果.对于待考察的特征矢量,通过 Viterbi 对准,可以得到有监督的正确状态标记;再经过无监督的 Viterbi 识别过程,得到识别状态标记,如果该帧的识别状态标记与正确状态标记不同,就认为该

帧属于“识别状态”的混淆集合.

考虑到前一种混淆数据采集方式只考虑单帧的“距离”, 可能存在某些特征帧, 尽管它们可能与错误状态中心距离较近, 但其所属的词句在识别中并不和那些状态描述的词句混淆, 那么采集这样的数据对提高识别率就没有意义. 而对于那些易混淆的词句, 它们的特征帧都应该在考虑之列. 而后一种混淆数据采集方法正与这种认识一致, 可以先根据多候选结果, 或是识别的似然度结果, 得到易混淆的词条, 再从该此词条的特征序列中得到需要的混淆集合, 因此后一种方式更适用于语音识别的混淆数据采集.

对所有训练语音库中的所有词条进行混淆数据采集, 详细包括 3 个步骤: (1) 对每条语音根据其发音标注进行 Viterbi 对准, 得到特征序列中的每帧特征的正确状态标注与似然度得分. (2) 对该词条进行 Viterbi 识别, 得到前 N 选识别结果与相应状态标注, 检查前 N 选的识别状态标注; 或者考虑前 N 选中与正确识别结果的似然度差值小于一定阈值的结果的识别状态标注. 凡是被标注成与其对应正确状态不同的特征矢量, 都被确认为易混淆特征数据, 并被采集到识别标注状态的混淆集合中. (3) 所有训练词条经过(1), (2)步处理后, 就得到了全部模型状态的混淆数据集合. 在进行线性区分分析时, 它们作为各个状态对应的混淆类别使用, 即需要和正确状态区分开的类别.

图 2 给出了汉语数码状态混淆集合的采集结果统计分布图, 图中显示了一些声母模型和韵母模型的状态混淆集合的样本来源. 这些混淆数据是直接从识别结果的前三选中采集到的. 图 2 中每个图上方的音标标记表示该图对应的是哪个声母或韵母. 图中各个部分表示这部分混淆数据来源于哪些声母和韵母, 大小显示了该部分在所有混淆数据中的比例. 所有图中都有一片没有标出混淆来源, 凡是提供混淆数据比例低于 5% 的来源都归并到这一片中.

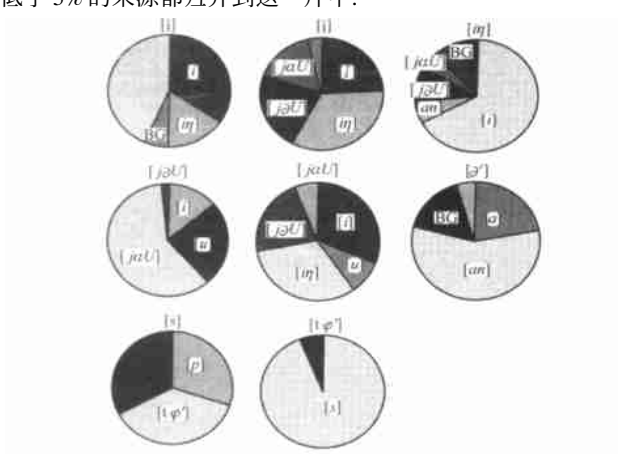

图 2　声母和韵母模型状态的混淆集合情况

从图 2 可以总结汉语数码的混淆情况. 不同的音素主要和与其发音相近的音素混淆, 这和汉语数码识别的混淆情况是一致的. 例如, 正是由于[i], [ɹ], [əʊ] 和[jaU] 之间存在的混淆, 造成了数字“0”,“A”,“6”和“9”之间的识别错误.

对所有模型状态的混淆集合收集完成之后, 就可以利用这些混淆数据集合, 进行基于状态的线性区分分析, 通过状态特定的线性区分变换, 将原始特征空间变换成新的空间下的特征. 这些变换后的特征对该状态下的易混语音数据有更强的鉴别能力. 在变换特征空间中进行样本识别, 使识别模型的整体鉴别能力得到很大提高.

3.3　状态特定的线性区分变换

从训练语音库可以得到各个状态的特征样本集合和相应混淆类的特征样本集合, 线性区分分析的类内和类间散度矩阵定义如下:

$$\boldsymbol{S}_w = \frac{1}{N_s}\sum_{i=1}^{N_s}(\boldsymbol{x}_{s,i}-\boldsymbol{m}_s)(\boldsymbol{x}_{s,i}-\boldsymbol{m}_s)^T \quad \boldsymbol{x}_{s,i}\in \text{状态 } S \tag{1}$$

$$\boldsymbol{S}_b = \frac{1}{N_{\bar{s}}}\sum_{i=1}^{N_{\bar{s}}}(\boldsymbol{x}_{\bar{s},i}-\boldsymbol{m}_s)(\boldsymbol{x}_{\bar{s},j}-\boldsymbol{m}_s)^T \quad \boldsymbol{x}_{\bar{s},i}\in \text{混淆类 } \bar{s} \tag{2}$$

等式(1)与(2)中, $\boldsymbol{m}_s = \frac{1}{N_s}\sum_{l}^{N_s}\boldsymbol{x}_{s,l}$, $\boldsymbol{x}_{s,i}$是状态s 的特征矢量样本, $\boldsymbol{x}_{\bar{s},i}$是状态 s 的混淆类的特征矢量样本, N_s 和$N_{\bar{s}}$ 分别是状态 s 和其混淆类$\bar{s}$ 的样本总数. 线性区分分析通过一个线性变换矩阵 $\boldsymbol{T}_{m\times n}$将 n 维的原始特征矢量$\boldsymbol{x}$ 变换成$\boldsymbol{m}$ 维$(m<n)$特征矢量 $\boldsymbol{y}$, $\boldsymbol{y}=\boldsymbol{Tx}$. 通过特征空间的线性变换 $\boldsymbol{T}$, 同时将类内和类间散度两个矩阵对角化:

$$\boldsymbol{T}\boldsymbol{S}_w\boldsymbol{T}^T = \boldsymbol{I}, \text{并且 } \boldsymbol{T}\boldsymbol{S}_b\boldsymbol{T}^T = \Gamma \tag{3}$$

其中, $\boldsymbol{I}$ 是单位阵, Γ 是对角阵.

上述变换的过程可详细分解成 4 个步骤, 其中主要包括特征空间的旋转, 特征数值的比例缩放和特征空间的降维等.

第一步, 利用类内散度矩阵 $\boldsymbol{S}_w$ 的特征矢量矩阵 $\boldsymbol{R}_1$ 对角化 $\boldsymbol{S}_w$, 即:

$$\boldsymbol{R}_1^T\boldsymbol{S}_w\boldsymbol{R}_1 = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) \tag{4}$$

这里 λ_i 是特征矢量矩阵 $\boldsymbol{R}_1$ 的各列特征矢量相应的特征值.

第二步, 对旋转后“椭球”的各个正交轴进行数值比例缩放, 变换成“单位圆球”, 变换式为:

$$\Lambda^{-1/2}\boldsymbol{R}_1^T\boldsymbol{S}_w\boldsymbol{R}_1\Lambda^{-1/2} = \boldsymbol{I} \tag{5}$$

其中 $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$. 同样, 类间散度矩阵 $\boldsymbol{S}_b$ 也要投影到新的变换空间中, 即:

$$\boldsymbol{S}'_b = \Lambda^{-1/2}\boldsymbol{R}_1^T\boldsymbol{S}_b\boldsymbol{R}_1\Lambda^{-1/2} \tag{6}$$

第三步, 利用 $\boldsymbol{S}'_b$ 的特征矢量矩阵$\boldsymbol{R}_2$ 对角化 $\boldsymbol{S}'_b$, 即:

$$\boldsymbol{R}_2^T\boldsymbol{S}'_b\boldsymbol{R}_2 = \Gamma \tag{7}$$

在对角化过程中, 类内散度矩阵在该空间中仍保持为单位阵.

第四步, 将特征空间的维数从 n 维减小到m 维, 去掉 Γ 的特征值小于 1 的维数, 保留特征值大于 1 的维数, 用 $\boldsymbol{F}_{n\times m}$ 表示降维变换. 经上述 4 步变换, 就实现了类内和内间散度矩阵的对角化, 总的变换矩阵:

$$\boldsymbol{T}^T = \boldsymbol{R}_1\Lambda^{-1/2}\boldsymbol{R}_2\boldsymbol{F} \tag{8}$$

原始特征空间变换到新的特征空间后, 相应状态的类内特征矢量的各个分量线性无关, 满协方差阵变换为对角协方差阵, 且各分量的方差都是 1, 所以只有变换后在混淆类样本集合中方差大于 1 的特征分量才具有区分状态类别和其对应混淆类别的能力, 方差小于 1 的特征分量难以区分这两个类别, 可以舍弃.

实验中发现某些状态的混淆类样本数量过少, 比如声母塞音[p]的状态. 此时对混淆类样本分布估计很不准确, 因此事先预定 200 帧作为混淆类样本数量阈值, 只有当混淆类的样本总数大于该值时, 才可以使用方程式(2)计算类间散度矩阵, 否则, 采用传统的线性区分分析定义的类间散度矩阵作为 $\boldsymbol{S}_b$ 的估值:

$$\boldsymbol{S}_b = \frac{1}{K}\sum_{i=1}^{K}(\boldsymbol{m}_i - \bar{\boldsymbol{m}})(\boldsymbol{m}_i - \bar{\boldsymbol{m}})^T \tag{9}$$

在方程(9)中 $\boldsymbol{m}_i$ 表示各个状态高斯分布的均值, $\bar{\boldsymbol{m}}$ 是所有模型状态的样本集合的均值, K 是状态总数.

3.4 变换空间中的高斯概率分布函数的计算

对所有隐含马尔科夫模型的状态进行线性区分分析后, 每个状态 S 都有一个线性变换 $\boldsymbol{T}(s)$, 将 n 维特征矢量变换成 m 维特征矢量 $\boldsymbol{y} = \boldsymbol{T}(s)\boldsymbol{x}$. 原始特征空间的高斯状态分布的均值和协方差阵分别变换成:

$$\mu_t = \boldsymbol{F}^T\boldsymbol{R}_2^T\Lambda^{-1/2}\boldsymbol{R}_1^T\mu \tag{10}$$

$$\Sigma_t = \boldsymbol{F}^T\boldsymbol{R}_2^T\Lambda^{-1/2}\boldsymbol{R}_1^T\Sigma\boldsymbol{R}_1\Lambda^{-1/2}\boldsymbol{R}_2\boldsymbol{F} \tag{11}$$

由于状态的协方差阵 Σ 就是 $\boldsymbol{R}_1\Lambda\boldsymbol{R}_1^T$, 并且 $\boldsymbol{R}_1$ 和 $\boldsymbol{R}_2$ 都是正交矩阵, 因此 Σ_t 等于单位矩阵:

$$\Sigma_t = \boldsymbol{I} \tag{12}$$

由方程(10)和(11)可得: 在变换空间中, 状态 s 对特征量 $\boldsymbol{y}$ 的输出概率分布函数为:

$$\begin{aligned} p(\boldsymbol{y}) &= \frac{1}{(2\pi)^{m/2}|\Sigma_i|^{1/2}}\exp\left\{-\frac{1}{2}[(\boldsymbol{y}-\mu_t)'\Sigma_i^{-1}(\boldsymbol{y}-\mu_t)]\right\} \\ &= \frac{1}{(2\pi)^{m/2}}\exp\left\{-\frac{1}{2}\|\boldsymbol{F}^T\boldsymbol{R}_2^T\Lambda^{-1/2}\boldsymbol{R}_1^T(\boldsymbol{x}-\mu)\|^2\right\} \\ &= \frac{1}{(2\pi)^{m/2}}\exp\left\{-\frac{1}{2}\|\boldsymbol{T}(s)(\boldsymbol{x}-\mu)\|^2\right\} \end{aligned} \tag{13}$$

在 Viterbi 搜索过程中, 对各状态似然度的计算可以在变换空间中直接进行. 该状态对特征矢量的输出概率分布函数, 用式(13)代替原始的高斯输出分布函数即可.

4 实验结果

为了详细比较基于状态特定的线性区分分析的结果, 在实验中分别采用了前 N 选阈值和似然度阈值两种混淆数据采集方法, 同时改变音素 HMM 模型的状态个数, 在不同状态数的条件下, 分别采集混淆数据. 通过这个实验, 可以比较线性区分分析对状态数不同的模型的区分性能的影响.

图 3 给出不同测试条件下的误识率变化曲线. 实验结果表明, 前 N 选混淆数据采集方法和似然度阈值采集, 线性区分分析的效果相近. 实验中音素模型的状态数从 1 增加到 4, 无论采用何种混淆数据采集方法, 严格阈值情况下得到的区分识别模型的识别性能都优于松弛阈值情况下的区分模型, 这说明每个状态采集到的混淆数据越集中在该状态和其混淆类别的样本"交遇区", 线性区分分析对这两类样本的区分性能越好. 此外从图中误识率随变换特征空间维数变化的曲线还可以看出, 不同模型达到最低误识率的混淆数据采集方法和确定的变换特征维数都不同, 最优的组合只能通过实验确定.

为了更清楚地给出状态特定线性区分分析对降低误识率作用, 表 2 给出了不同模型经变换后达到最低误识率的变换情况和误识率相对降低比率. 表 2 的第二列说明, 即使是所有半音节模型只有 1 个状态, 此时识别模型非常粗糙, 系统的识别率很低, 约为 82.61%, 经过状态特定的线性区分分析后, 识别率也提高到 97.5%. 当半音节模型的状态数增加到合适的数量后, 达到了> 99% 的识别率. 表 2 的结果说明, 基于状态特定混淆数据区分的线性区分分析能有效降低误识率, 当模型的状态数较多时, 系统识别率能够提高到 99% 以上.

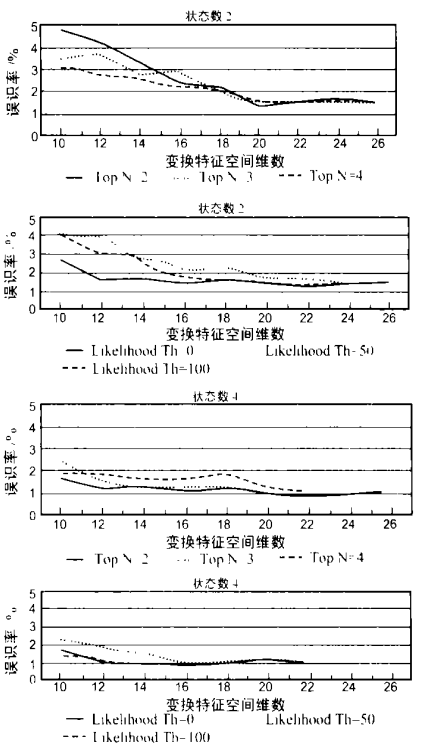

图 3 不同模型在多种线性区分分析方法下的误识率比较

表 2 状态数不同半音节模型线性变换前后的误识率比较

状态数	1	2	3	4
变换前误识率	17.39%	4.32%	2.95%	1.82%
变换后误识率	2.5%	1.02%	0.91%	0.8%
误识率下降比率	85.6%	76.4%	69.2%	56.0%
变换特征维数	24	18	16	22
混淆数据采集方法	似然度 100	似然度 150	前 2 选	前 2 选

在最后的测试实验中, 根据不同音素模型对应的发音平均时长的不同, 适当调整对应 HMM 模型的状态数, 表 3 分别给出了 15 个音素模型的最终采用的状态数. 表 4 给出了经线性区分变换后, 达到最低误识率结果. 此时混淆数据采集方法采用前 3 选方法, 变换的特征维数为 22 维.

表 3 15 个音素 HMM 模型状态数

模型	[l]	[p]	[s]	[tɕʰ]	[tʂ]
状态数	3	2	3	4	3
模型	[i]	[ɑʳ]	[ɿ]	[jɔU]	[jɑU]
状态数	4	7	5	6	8
模型	[ĩ]	[a]	[u]	[.]	[an]
状态数	4	5	7	3	5

表 4　线性区分分析变换的识别率比较

变换前识别率	变换后识别率	前 2 选	前 3 选	前 4 选
97. 16%	99. 32%	99. 77%	99. 89%	100%

以上实验说明,经过状态特定的线性区分分析变换后,模型对混淆数码的识别率得到相当的提高. 鉴于这种线性区分变换是针对音素模型的每个状态和其混淆集合进行的,可以说在变换特征空间内,汉语数码中常见的音素混淆情况得到了相当的改善. 从实验结果看,模型对相近元音[i] 和[î] 之间的鉴别特性得到很大改善,并且这两个音与其它包涵有相似音素的半音节[ɚl] ,[jɔ̂U] 和[jɑU] 之间的区分性也有明显提高;同样,[ə̂r] 和[a] ,以及[an] 之间的区分性有了提高;擦音[tɕʻ] 和[s] 之间的区分性也得到改进. 模型对这些音素的鉴别特性的提高,反映为对混淆数码集合{ 0, A, 6, 9} ,{ 2, 8} 和{7, 4} 的识别性能的改善,最终有效提高了汉语数码的识别率.

5　结论

本文对汉语数码识别中的混淆问题进行了细致分析,在声学模型的状态层次应用线性区分分析方法,利用基于状态混淆集合区分的线性变换,有效地提高了识别模型对易混数码语音之间的鉴别能力,使孤立汉语数码的识别率达到 99. 32%. 这是目前有关文献报导的最好水平.

本文采用的线性区分分析实际上有一个假设前提条件,即待区分模式类的类内协方差矩阵相同. 只有在该假设条件下,对角化线性变换才是可分离性判据最优化的变换. 当模式类的分布不同时,如何求解最优变换仍是研究课题之一. 参考文献[11, 12] 给出了两种无需等协方差假设,进行线性区分变换的方法. 关于如何在模型不同层次上采集混淆数据问题,在参考文献[4~ 7, 13] 中给出了其它方法. 基于最大似然估计原理,参考文献[11, 14] 采用了最优线性区分变换和隐含马尔科夫模型参数训练相结合的方法. 此外,基于判决树方法,对一定语音集合内状态的线性区分变换矩阵进行捆绑训练,可在保证性能的前提下有效减少区分变换参数所需的存储量和计算量. 这些将是我们在下一步研究中需要考虑的问题.

附录:

数字	0	1	2	3	4	5	6	7	8	9	A
拼音	ling	yi	er	san	si	wu	liu	qi	ba	jiu	yao
国际音标	[lɚl]	[i]	[ə̂r]	[san]	[sî]	[u]	[ljɔ̂U]	[tɕʻ i]	[pa]	[tɕʻjɔ̂U]	[jɑU]
声母	[l]	[p]	[s]	[tɕʻ]	[tɕʻ]	[.]					
韵母	[î]	[i]	[ə̂r]	[u]	[a]	[an]	[ɚl]	[jɔ̂U]	[jɑU]		

参考文献:

[1] K Fukunaga. Introduction to Statistical Pattern Recognition [M] . NY: Academic Press, 1990.

[2] M J Junt, C Lefebvre. Speaker dependent and independent speech recognition experiments with an auditory model [A] . Proc. ICASSP [C] . USA: ICASSP, 1988. 215- 218.

[3] M J Hunt, C Lefebvre. A comparison of several acoustic representation for speech recognition with degraded and undegraded speech [A] . Proc. ICASSP [C] . USA: ICASSP, 1989. 262- 265.

[4] C J Leggetter. Improved acoustic modelling for HMMs using linear transformations [D] . UK: Cambridge, 1995

[5] R Haeb-umbach, H Ney. Linear discriminant analysis for improved large vocabulary continuous speech recognition [A] . Proc. ICASSP [C] . USA: ICASSP, 1992: 13- 16.

[6] E S Parris, M J Carey. Estimating linear discriminant parameters for ccintinuous density hidden Markov models [A] . Proc. ICSLP [C] . USA: ICASSP, 1994. 215- 218.

[7] P C Woodland, D R Cole. Optimising hidden Markov models using discriminative output distributions [A] . Proc. ICASSP [C] . USA: ICASSP, 1991.

[8] Duda Richard O, Hart Peter E. Pattern Classification and Scene Analysis [M] . New York: John Wiley & Sons, Inc. , 1973.

[9] E L Bocchieri, G R Doddington. Frame specific statistical features for speaker independent speech recognition [J] . IEEE trans. ASSP, 1986, 34(4) : 755- 764.

[10] G R Doddington. Phonetically sensitive discriminants for improved speech recognition [A] . Proc. ICASSP [C] . USA: ICASSP, 1989. 556 - 559.

[11] Nagendra Kumar, Andreas G. Andreou, Heteroscedastic discriminant analysis and reduced rand HMMs for improved speech recognition [J] . Speech communication, 1998, 26: 283- 297.

[12] Marco Loog, Reinhold Haeb-umbach. Multi-class linear dimension reduction by generalized fisher criteria [A] . Proc. ICSLP [C] . USA: ICASSP, 2000.

[13] T Hastie, R Tibshirani. Discriminant analysis by Gaussian mixtures [J] . Journal of the Royal Statistical Society, series b. 1996, 58: 155- 176.

[14] George Saon, Mukund Padmanab, et al. Maximum likehood discriminant feature spaces [A] . Proc. ICASSP [C] . USA: ICASSP. 2000.

作者简介:

史媛媛　女, 1974 年出生于河南省洛阳市, 1997 年于北京航空航天大学获得学士学位,现为清华大学电子工程系博士研究生,主要研究方向为语音识别算法与应用,以及语音识别技术的 ASIC 实现.

刘　加　男, 1954 年生于福建, 1983 年获得清华大学电子工程系无线电技术专业学士学位, 1986 年获得清华大学电子工程系通信与电子系统硕士学位, 1990 年获得清华大学电子工程系通信与电子系统博士学位, 1990 年至 1992 年中科院遥感卫星地面站从事美国 6 号陆地卫星图像处理系统开发工作, 1992 年至 1994 年在英国剑桥大学作博士后,从事语音识别系统工作,现任清华大学电子系副教授,中国电子学会高级会员,美国 IEEE 会员,目前研究方向包括语音识别,语音合成,语音编码,语音识别专用芯片设计,多传感器融合技术,以及多媒体数字通信系统.

刘润生　男, 1933 年生于河北,清华大学电子系教授,常年从事数字电路, IC 设计,电子线路 CAD 和语音信号处理的教学与研究工作.