一种层次化空间分析方法在语种识别系统中的应用*

常振超，刘　斌，石远超，张兴明，杨镇西，张　丽
（解放军信息工程大学 信息工程学院，郑州 450002）

摘　要： 在针对电话语音的自动语种识别系统中，训练和测试语料之间存在不同说话人、信道等因素差异带来的不匹配，是影响识别性能提高的关键因素。为了消除此类影响，提出一种层次化空间分析方法，首先对前端部分 MFCC + SDC 特征进行 HLDA（异方差线性判别分析），增大了语种各个类的类间差异；然后对经自适应得到含有冗余信息的 GSV 进行 PCA 特征选择，有效地去除了信道等冗余信息的干扰。实验结果表明，此方法能有效消除信道等噪声影响，从而提升了原有系统的识别性能。

关键词： 语种识别；层次化；空间分析；冗余信息
中图分类号： TN912.34　　**文献标志码：** A　　**文章编号：** 1001-3695(2012)10-3651-04
doi:10.3969/j.issn.1001-3695.2012.10.013

Language recognition method based on hierarchical space analysis

CHANG Zhen-chao, LIU Bin, SHI Yuan-chao, ZHANG Xing-ming, YANG Zhen-xi, ZHANG Li
(*Institute of Information Engineering, The PLA Information Engineering University, Zhengzhou* 450002, *China*)

Abstract: In automatic spoken language recognition system on telephone conversation speech, differences between train and test utterances on channels, gender and speakers are the key factor of improving the performance of the system. This paper proposed a hierarchical space analysis method. Firstly, it mapped the front-end cepstral features of SDC into the HDLA space, aiming at increasing the discriminability between different languages. Secondly, it selected the characters of adaptive GMM super vector by the method of PCA, which eliminated the influences of different channels, speakers and so on. Experiment results indicate that this method is better for improving the system's performance than the original baseline system.

Key words: language recognition; hierarchical; space analysis; redundant information

0　引言

语种识别通过对给定的一段语音信号进行处理，识别其所属语言的种类，往往作为语音处理的一个前端处理技术，在信息检索和安全领域都有着重要的应用。基于高斯混合模型超矢量—支持向量机（Gaussian mixture model super vector-support vector machine，GSV-SVM）[1] 的语种识别系统因性能优良、可移植性强、训练语料无须音素标注且算法复杂度低，是当前研究热点。在针对电话语音的语种识别系统中，数据背景噪声较大，同时受到不同说话人、通话信道以及不同说话人的停顿、笑声和咳嗽声等非语音因素的干扰严重，如何在语种识别系统中除去这些因素的影响，是提高语种识别性能的关键。

针对上述问题，多家语种识别机构开始采用子空间分析方法如主分量分析（principal component analysis，PCA）、异方差线性判别分析（heteroscedastic linear discriminant analysis，HLDA）等，通过寻求有效的子空间表述以达到去除非语种因素干扰的目的，这方面的研究工作在最近几年成为热点。布尔诺科技大学的 Matejka[2] 在其 2008 年的博士毕业论文中将 HLDA 应用在基于 MMI 准则下的 GMM-UBM 的语种识别系统中前端的 56 维特征参数上，发现经 HLDA 投影后，能有效提高系统的识别性能，同时达到降维的目的，但其并没有将此类方法应用在基于 GSV-SVM 的语种识别系统中。宋彦等人[3] 于 2010 年在基于 GSV-SVM 的语种识别系统中对 GSV 超矢量应用了核 PCA 进行子空间分析，提出核 PCA 子空间分析是因子分析的一种，能够对 GSV 空间进行噪声空间的估计，有效去除了信道噪声的影响，达到了提升识别性能的效果。在 2011 年的 NIST LRE 公布的参赛文档中，i-vector 系统已经成为各家参赛机构广泛采用的有效的语种识别系统。麻省理工学院在其公布的 2011 NIST LRE 技术手册 Mitll_lre11_descriptio[4] 中提到，在 i-vector[5] 系统中采用了首先使用 HLDA 算法进行去除语种间的差异，然后借用因子分析[6] 中子空间估计的方法，基于本征特征信息捕获最主要的信息变量，然后经由特征变换用低维的有效空间（iVectors）来描述语种和信道信息，并在此空间下训练和识别语种，取得了较好的识别效果。

基于以上各个系统的分析，本文借鉴子空间分析思想，提出一种基于空间变换的分析方法。首先在 56 维的语音特征参数上进行 HLDA 映射，增加了各个语种类间的区分性，减少了输

收稿日期： 2012-03-08；**修回日期：** 2012-04-19　　**基金项目：** 国家“863”计划重点资助项目（2008AA011002）

作者简介： 常振超（1987-），男，河北邯郸人，硕士，主要研究方向为语种识别和 FPGA 设计（changzhenchao1987@126.com）；刘斌（1986-），男，河南郑州人，硕士，主要研究方向为通信与信息系统；石远超（1984-），男，河南濮阳人，硕士，主要研究方向为通信与信息系统；张兴明（1963-），男，河南新乡人，教授，硕导，主要研究方向为通信与信息系统；杨镇西（1971-），男，山东鄄城人，工程师，主要研究方向为语种识别和系统级芯片（SoC）设计；张丽（1982-），女，河南新乡人，讲师，主要研究方向为语种识别和系统级芯片（SoC）设计.

入特征的维数,然后对经自适应得到的 GSV 超矢量进行核主成分分析(KPCA);在空间分析中,对系统信道噪声影响进行了一定程度上的抑制,同时减少了后端 SVM 模块分类判决模块输入的特征矢量的维数,从而提高了其训练和测试的实时性。

1 基于 GSV-SVM 的语种识别系统简介

基于 GSV-SVM 的语种识别系统分为前端特征提取和后端建模分类两个部分。其中,特征提取部分包括语音信号预处理和特征参数提取。预处理包括语音信号数字化(8 kHz 采样,16 bit 量化)、预加重(系数为 0.97)、加汉明窗(窗长 25 ms)和分帧(帧长 25 ms,帧移 10 ms)。特征提取时,首先提取 7 维的美尔频率倒谱系数(Mel-frequency cepstral coefficients,MFCC),并采用半升正弦倒谱提升和倒谱均值归一化(cepstral mean normalization, CMN)技术抑制特征域噪声;然后将 7 维 MFCC 与其 7-1-3-7 的滑动差分倒谱(shifted delta cepstral,SDC)[1]构成 56 维特征参数;最后对这 56 维参数进行能量归一化并利用语音激活检测(voice activity detection,VAD)技术剔除静音帧。后端建模分类部分可分为训练和测试两个阶段。训练阶段在 UBM 上通过 MAP 自适应得到各训练语音的 GSV 用于训练 SVM 模型;识别阶段采用同样的方法得到各测试语音的 GSV,然后经训练好的 SVM 进行识别,输出结果。GSV-SVM 语种识别系统如图 1 所示。

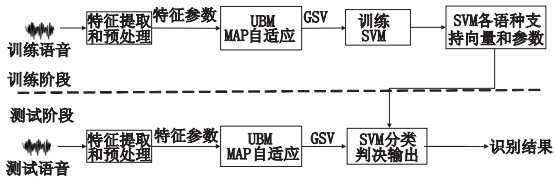

图1 GSV-SVM语种识别系统

1.1 GSV 构成

GMM 是用多个单高斯分布的线性组合来描述帧特征在特征空间的分布,即

$$g(\boldsymbol{x}) = \sum_{i=1}^{M} w_i N(\boldsymbol{x}, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \tag{1}$$

其中:$\boldsymbol{x}$ 为语音帧声学特征向量,M 为高斯混合数,w_i 为混合权重,$\boldsymbol{\mu}_i$ 和 $\boldsymbol{\Sigma}_i$ 为第 i 个高斯混合成分的均值向量和协方差矩阵。首先用训练数据通过期望最大化算法(expectation maximum, EM)[1]得到一个通用背景模型(universal background model, UBM)。每一个训练和测试的语句,通过最大后验概率(maximum a posteriori,MAP)准则从 UBM 中自适应得到各自对应的 GMM 模型;然后将自适应得到的各高斯混合成分的均值向量按顺序排列起来即构成超矢量(GSV)。

1.2 SVM 的训练和测试

SVM 是一种应用广泛的机器学习方法。在二分类问题中,给出样本 $\{\boldsymbol{x}_i, y_i\}$,$i = 1, 2, \cdots, N$,$\boldsymbol{x}_i \in \mathrm{R}^D$ 为 D 维的特征向量,$y_i \in \{+1, -1\}$ 为类别标签,其分类判决函数表示为特征向量内积的形式。

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_i (\boldsymbol{x} \cdot \boldsymbol{x}_i) + b \tag{2}$$

其中:$\sum_{i=1}^{N} \alpha_i = 0$,$\alpha_i < 0$ 表示 $\boldsymbol{x}_i$ 为负类支持向量,$\alpha_i > 0$ 表示 $\boldsymbol{x}_i$ 为正类支持向量;支持向量是处于分类边界上的样本点。

对于非线性的问题,通常采用核函数将输入特征向量(即 GSV)非线性地映射到高维空间,当做线性问题处理。核函数形式为 $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \varphi(\boldsymbol{x}_i) \cdot \varphi(\boldsymbol{x}_j)$,这样在高维空间只需进行内积运算即可,判决函数转换为

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_i k(\boldsymbol{x}, \boldsymbol{x}_i) + b = \sum_{i=1}^{N} \alpha_i \varphi(\boldsymbol{x}) \cdot \varphi(\boldsymbol{x}_i) + b \tag{3}$$

本文 SVM 的核函数采用度量 GMM 距离的 Kullback-Leibler 核函数(K-L 核)[1],然后利用 SVM 的线性核函数在 K-L 空间进行训练和识别。

2 层次化空间分析方法在语种识别系统中的应用

基于 GSV-SVM 的语种识别系统中,GSV 作为一段语音的特征表示,包含了说话人、语种和信道等多方面的因素。而在 GMM-SVM 中的序列核函数 $K(\cdot, \cdot)$,则是从 GSV 之间的距离测度中推导得到。在给定语种类别标号之后,可将均值超矢量之间差值进行建模,如下所示[3]:

$$\boldsymbol{d}_{ij} = \boldsymbol{u'}_i - \boldsymbol{u'}_j = \begin{cases} O + I + N & \boldsymbol{u'}_i, \boldsymbol{u'}_j \in \text{不同语种} \\ I + N & \boldsymbol{u'}_i, \boldsymbol{u'}_j \in \text{相同语种} \end{cases} \tag{4}$$

其中:I 为语种内部差异,表示同一语种中因说话内容和说话人不同等产生的差异;噪声影响 N 表示不同语音段存在的信道背景噪声差异以及其他一些人声非语音(铃音、咳嗽和停顿等)的因素;O 为语种之间差异,GSV 中存在大量的对于区分语种无效的冗余信息。如果能在距离度量中去除 I 和 N 的影响,可有效地提高语种识别性能。

2.1 异方差线性判别分析

异方差线性判别分析(HLDA)可以用来线性减少特征之间的相关性同时实现降维的目的[2]。对于 HLDA,用来产生变换矩阵的特征矢量要属于同一个类。HLDA 实现降维主要是通过保留能够有效进行分类的有用维,同时去除干扰分类的无用维来实现的。

对于某一特定的类,HLDA 实现的是去除特征相关性的最优解,文献[7]给出了详细证明。降维是通过 HLDA 变换得到的,通过 HLDA 变换将 n 维特征映射到 p 维,其中 $p < n$。设 HLDA 的变换为维数 $n \times n$ 的矩阵 $\boldsymbol{A}$,取其前 p 行($p < n$)作为降维矩阵,本文中 $\boldsymbol{A}$ 的估计使用文献[8]中迭代算法,每行的重估公式为

$$\hat{a}_k = c_k G^{(k)-1} \sqrt{\frac{T}{c_k G^{(k)-1} c_k^{\mathrm{T}}}} \tag{5}$$

其中:c_i 为伴随矩阵 $\boldsymbol{C} = |\boldsymbol{A}| \boldsymbol{A}^{-1}$ 的第 i 行($\boldsymbol{A}$ 使用当前的估计值):

$$G^{(k)} = \begin{cases} \sum_{j=1}^{J} \dfrac{\gamma_j}{a_k \hat{\boldsymbol{\Sigma}}^{(j)} a_k^{\mathrm{T}}} \hat{\boldsymbol{\Sigma}}^{(j)} & k \leqslant p \\ \dfrac{T}{a_k \hat{\boldsymbol{\Sigma}} a_k^{\mathrm{T}}} \hat{\boldsymbol{\Sigma}} & k > p \end{cases} \tag{6}$$

其中:$\hat{\boldsymbol{\Sigma}}$ 和 $\hat{\boldsymbol{\Sigma}}^{(j)}$ 分别为全局协方差矩阵和第 j 类的协方差矩阵,γ_j 为训练特征矢量属于第 j 类的概率得分,T 为训练特征矢量的总数。

因此,经过 HLDA 降维后混元 j 的均值 $\hat{\boldsymbol{\mu}}_j^{\mathrm{HLDA}}$ 和协方差向量 $\hat{\boldsymbol{\Sigma}}_j^{\mathrm{HLDA}}$ 为

$$\hat{\boldsymbol{\mu}}_j^{\mathrm{HLDA}} = \boldsymbol{A}_p \hat{\boldsymbol{\mu}}_j \tag{7}$$

$$\hat{\boldsymbol{\Sigma}}_j^{\mathrm{HLDA}} = \mathrm{diag}(\boldsymbol{A}_p, \hat{\boldsymbol{\Sigma}}_j \boldsymbol{A}_p^{\mathrm{T}}) \tag{8}$$

其中:$\boldsymbol{A}_p$ 为变换矩阵的 $\boldsymbol{A}$ 的前 p 行。

2.2　超向量的核主成分分析

针对任一训练集 $X=\{(\boldsymbol{x}_i,l_i)\}$,$i\in\{1,2,\cdots,p\}$,其中 $\boldsymbol{x}_i$ 为归一化后的 GSV 均值超矢量,$\boldsymbol{x}_i\in\mathrm{R}^D$,$D$ 为该样本的特征参数;l_i 为选取样本的语种类别,且 $l_i\in\{1,2,\cdots,L\}$。与之对应的协方差矩阵 $\boldsymbol{C}$ 可以由式(9)计算而得出。

$$\boldsymbol{C}=\sum_{i=1}^{p}\sum_{j=1}^{p}(\boldsymbol{x}_i-\boldsymbol{x}_j)(\boldsymbol{x}_i-\boldsymbol{x}_j)^{\mathrm{T}} \tag{9}$$

由式(9)分析可知,原始训练集 X 的协方差矩阵可以用相应的差值超矢量 $\boldsymbol{d}_{ij}=\boldsymbol{x}_i-\boldsymbol{x}_j$,$(i=1,2,\cdots,p;j=1,2,\cdots,p)$ 的乘积形式进行描述。对该协方差矩阵的 PCA 分析,可以理解为此协方差矩阵刻画了训练语句中任意两个超矢量之间的差异分布情况,此差异主要体现为语种之间的差异,而共存于整个特征矢量中的电话信道噪声则会被消除掉。在对合适数目的本征向量进行选取后,最终得到的 PCA 子空间中保留了主要的语种间的差异分量;而噪声影响,如信道差异和其他人声非语音的部分则被有效地抑制。

在文献[9]中 Bishop 指出针对 $D\times D$ 维矩阵,在对其进行本征分解后所需要的运算复杂度为 $O(D^3)$,当对均值矢量空间采用核 PCA 的方法进行分析时,具体计算如下:

$$\begin{aligned}&\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{u}_i=\lambda_i\boldsymbol{X}^{\mathrm{T}}\boldsymbol{u}_i\\&\boldsymbol{K}\boldsymbol{v}_i=\lambda_i\boldsymbol{v}_i\end{aligned} \tag{10}$$

其中:$\boldsymbol{K}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 为线性 SVM 的核矩阵,$\boldsymbol{v}_i=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{u}_i$ 为矩阵的本征向量。上式表明 PCA 分析本质是在线性核空间中寻找有效的子空间表示,其运算复杂度与训练样本数目相关,为 $O(P^3)$。本文采用文献[10]中的期望最大化 PCA(expectation maximization PCA,EMPCA)的方法来实现算法,此算法可以对高维特征且数目较大的数据较快地计算出所用到的本征向量。对 GSV 进行的核 PCA 分析得到的是降低后的维数据,类似于一种简化的因子分析[3]。

2.3　层次化空间分析方法

结合之前的分析,本文提出子空间分析方法。首先,针对前端得到的 56 维语音特征参数相关性强的问题,对该特征进行 HLDA 映射,增加语种不同类别之间的区分性。然后,针对自适应得到的 GSV 维数高、相关性大的问题,使用核 PCA 对 GSV 进行特征选择,在子空间中去除信道环境噪声等的干扰,同时使得整体 GSV 的维数大大降低,提高了后端分类判决模块识别系统的实时性,并保证了其识别性能。

在训练阶段,需要训练 HLDA 和 PCA 的转换矩阵,同时需要用转换后的特征得到各个语种的 SVM 模型。此空间变换的特征维数约简算法的实现步骤如下:

a)在本文中,针对用于训练 HLDA 和 PCA 转换矩阵的语料,对其前端的 56 维语音特征参数进行 HLDA 分析,其中类的定义为某个语种的每一个高斯混合成分的权重,根据标准 GMM 进行训练时所提取得到每一特征矢量的概率得分为 $\gamma_j(t)$,通过此得分来判断该特征所属于的类别 j,用当前的得分与估计得到的第 j 类的协方差矩阵 $\hat{\sigma}_j$。

b)利用式(5)来重新估计新的变换矩阵 A,最后取其前 p 行作为最终特征的 HLDA 变换矩阵 A_p,并对高斯混元的均值和权重分别进行重估,将得到的新的特征用于后面的模型训练。

c)将降维后经自适应得到的 GMM 模型的均值和协方差分别取出,拼接成 GSV。得到带有语种标号的训练数据的超向量集合,$X=\{(\boldsymbol{x}_i,l(\boldsymbol{x}_i))\}$,$i\in\{1,2,\cdots,p\}$,其中 $\boldsymbol{x}_i$ 为归一化后的均值超矢量,$\boldsymbol{x}_i\in\mathrm{R}^D$,$D$ 为样本的特征参数;$l(x_i)$ 为样本的语种类别,$l(x_i)\in\{1,2,\cdots,L\}$。

d)得到由各个语种的 GSV 拼接成的特征向量,利用式(9)计算此超矢量矩阵的协方差矩阵 $\boldsymbol{C}$。

e)计算 $\boldsymbol{C}$ 的本征向量、优化的 PCA 子空间维数,得到最优转换矩阵 $\boldsymbol{A}$,同时去除信道等干扰噪声的影响。

f)得到用于训练各个语种的语料的特征超矢量,利用上述步骤得到的转换矩阵对其进行处理,得到降维后的特征用于各个语种的 SVM 模型参数的训练。

在测试阶段,针对测试语料提取的 56 维语音参数,由训练时得到的 HLDA 最佳参数和投影矩阵进行变换;同时,后端的 GSV 模块根据训练得到的转换矩阵进行测试语料的特征转换,再进行 SVM 的分类判决。

此空间变换的算法的简要实现流程如图 2 所示。

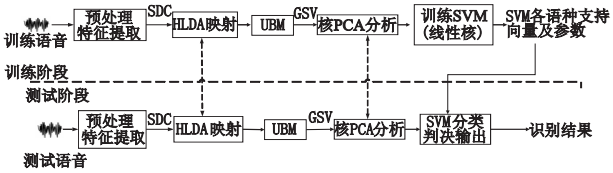

图2　层次化空间分析方法的GSV-SVM语种识别系统

3　语种识别实验结果及分析

3.1　实验设置及评价标准

语料库为实验室采集电话信道下的通话语音,采样频率为 8 kHz,并经过 16 bit 量化处理。语料库分为汉语普通话、英语和日语三个语种,共有 1 500 段 3 min 的语音、3 000 段 30 s 的语音和 1 500 段 10 s 的语音。30 s 语料中,每个语种有 1 000 段语音,男女各为 500 段;10 s 语料中,每个语种有 500 段语音,男女各为 250 段;3 min 长时语料中,每个语种有 500 段话音,男女各有 250 段。上述语音段均按照不同的说话人进行采集,各个语音段已经过处理,为单向通话语音,即每段语音仅含一个说话人。在本章的实验中,实验语料采用上述语料中的 3 000 段 30 s 话音和 1 500 段 10 s 语音进行基线系统的性能验证。

将语料库分为训练集和测试集两个部分。其中用于训练 UBM 的语料选择 30 s 时长的语料,挑选方式为:从每个语种的训练集中挑选 400 段语音(男女各 200 段),共 1 200 段用于训练 GMM-UBM 模型,混合数为 512(经验值);然后,从 30 s 中剩余的 1 800 段语料中挑选 400 段语音(男女各 200 段),共 1 200 段,作为训练 SVM 的语料。测试阶段的语料包括两种时长的测试集,一种是 30 s 时长,剩余的各个语种 200 段(男女各 100 段),共 600 段语料;另一种是上述的 1 500 段 10 s 时长的语料。

本文 SVM 的训练与识别部分,使用林智仁等人开发 LibSVM 工具包[11],这套工具包由于程序简洁、运用灵活,并且是开源的,易于扩展,是目前应用较多的 SVM 库。实验中特征分别采用当前主流的 MFCC + SDC 参数,其中 SDC 的参数(N, D, P, K)设为 7-1-3-7。对每帧语音加汉明窗,窗长取 25 ms,窗移 10 ms。语音特征提取之前进行话音激活检测(VAD),特征提取之后去除静音帧。GMM-UBM 混元数为 512,初始化用 LBG 算法,迭代 10 次。测试采用语种确认实验。例如进行汉语确认时,汉语是目标语种,英语和日语测试语音为非目标语种,对语种识别的结果 EER 进行表征,以此为准则测评语种识系统的性能。

3.2 实验结果及分析

实验 1 HLDA 参数的比较

对于基线系统 GSV-SVM 的 SDC 特征进行 HLDA 映射,使用 K-L 核作为实验中使用的核函数。如图 3 所示,经 HLDA 区分性投影后,识别性能得到了提高,表现为 EER 降低。

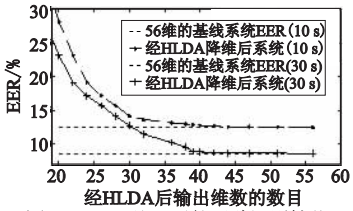

图3 HLDA处理后的语种识别性能

图 3 中,虚线表示为基线系统的 EER,其特征参数为 56 维,带标号的连线是经 HLDA 降维后的 EER。由图 3 可知,无论是 10 s 还是 30 s 的语音段,当 p 值选择小于 40 时,系统识别性能会急剧下降,表现为 EER 的值升高。究其原因是对于维数约简太多,从而对语种信息相关的特征损失太多,导致识别率下降。当 p 达到 40 左右时,系统识别率接近基线系统的识别率,且性能提升幅度逐渐趋于平缓。其原因是之前的 56 维特征含有冗余信息,当选择 40 维时有效地消除了冗余;但超过 40 维后性能效果提升不明显。在经过 HLDA 处理后仍选择维数为 56 时,发现此处理起到一定的提高语种识别性能的效果,原因是处理过程中拉大了不同语种间的距离,起到了增强语种识别效果的作用。根据实验结果及算法的保证识别性能条件下尽量提高实时性的目的,选取 $p=40$ 作为后续实验的参数。

实验 2 PCA 维数的选择

在确定好 HLDA 的参数选择为 40 维之后,将此引入系统中,并进一步对 GSV 超矢量进行 PCA 维数约简分析,实验结果如图 4 所示。

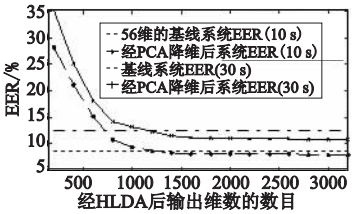

图4 PCA处理后的语种识别性能

在此实验中,对实验 1 中经 HLDA 降维后的特征获取其 GSV,并采用 PCA 特征转换方法对新得到的超向量空间进行特征变换分析,通过选择不同本征向量,得到不同的转换特征,并用此特征进行后端的分类判决。从实验中可以看出,当特征转换维数选择为 1 000 维左右时,系统的识别性能已经接近基线系统,分析可知特征变换有效消除了冗余信息,并增加了语种的区分性。随着维数进一步增加,识别性能较基线系统有所提升,其中 10 s 的语音片段性能提升较为明显,这是因为,当 10 s 语音经 UBM 模块自适应获取 GSV 时,由于语料较短,含有的信息量少,因此相对于 30 s 的语音段,更新的混合高斯较少,冗余信息也较大。从图 4 可以看出,当维数超过 1 400 维后,性能提升不明显,为减少运算量,本文中维数选择为 1 400 维。在运算量方面,对于整个系统而言,虽然此算法需要得到特征变换模块,且在特征进行变换时的运算量较大需要对特征进行一个 $1400\times20480(40\times512=20\ 180)$ 的矩阵相乘变换,但经过特征变换后输入给 SVM 的模块的维数为 1 400 维,这样大大减少了 SVM 模型的训练和测试时的运算量,避免了 SVM 训练时核空间的“维数灾难”。

实验 3 各个系统的识别性能比较

为评价本文所提算法的识别性能,共搭建了三个语种识别系统,分别是:未加任何补偿措施的 GSV-SVM 语种确认系统、基于 HLDA 降维后的 GSV-SVM 语种确认系统、基于精细化子空间分析的结合了 HLDA 和 PCA 分析的 GSV-SVM 语种确认系统。通过比对这三个系统的性能以验证本文所提方法的有效性,实验结果如表 1 所示。

表 1 各系统的 EER 对比

语音	语种识别系统 EER/%	
	基线系统	特征转换后
10 s	12.52	11.25
30 s	8.62	8.10

由表 1 可以看出,经过特征变换处理后的系统,无论是 10 s 还是 30 s 的语音,性能均有所提升,表现为 EER 的降低,且 10 s 时长下的性能提升要高于 30 s 的语音段,这也验证了之前的分析。

4 结束语

针对电话信道的语种识别系统存在的语种之外的干扰因素,本文在研究目前流行的子空间分析方法上提出了一种层次化空间分析方法。首先对前端得到的 SDC 特征进行 HLDA 分析,提高了语种的区分性,然后对经自适应得到含有冗余信息的高维超矢量空间进行 KPCA 特征选择,在子空间中去除信道等噪声影响。实验结果表明了该方法的有效性,进一步的研究工作是如何借助于最新的空间分析技术,寻找更为有效的空间分析方法,进一步消除语种无关的信息。

参考文献:

[1] CAMPBELL W M, CAMPBELL J P, REYNOLDS D A, *et al*. Support vector machines for speaker and language recognition[J]. Computer Speech & Language, 2006, 20(2-3): 210-229.

[2] MATEJKA P. Phonotactic and acoustic language recognition[D]. Brno: Brno University of Technology, 2008.

[3] 宋彦, 戴礼荣, 王仁华. 基于超向量子空间分析的自动语种识别方法[J]. 模式识别与人工智能, 2010, 23(2): 165-170.

[4] DEHAK N, McCREE A, REYNOLDS D, *et al*. MITLL 2011 language recognition evaluation system description[R]. Cambridge: Information Systems Technology Group, MIT Lincoln Laboratory, 2011.

[5] MARTINEZ G D, PLCHOT O, BURGET L, *et al*. Language recognition in iVectors space[C]//Proc of the 12th Annual Conference on International Speech Communication Association. 2011: 861-864.

[6] KENNY P. Joint factor analysis of speaker and session variability: theory and algorithms[EB/OL]. (2006-01-13). http://www.crim.ca/perso/patrick.kenny/FAtheory.polf.

[7] KUMAR N. Investigation of silicon-auditory models and generalization of linear discriminant analysis for improved speech recognition[D]. Baltimore: John Hopkins University, 1997.

[8] GALES M J. Semi-tied covariance matrices for hidden Markov models[J]. IEEE Trans on Speech and Audio Processing, 1999, 7(3): 272-281.

[9] BISHOP C M. Pattern recognition and machine learning[M]. New York: Springer, 2006.

[10] ROWIES S. EM algorithm for PCA and SPCA[C]//Proc of Conference on Advances in Neural Information Processing System. Cambridge: MIT Press, 1997: 626-632.

[11] http://www.csie.ntu.edu.tw/~cjlin/[EB/OL].