基于贝叶斯主成分分析的i-vector说话人确认方法

肜娅峰[1],陈 晨[1,2],陈德运[1,2],何勇军[1]
(1. 哈尔滨理工大学计算机科学与技术学院,黑龙江哈尔滨 150080;
2. 哈尔滨理工大学计算机科学与技术博士后流动站,黑龙江哈尔滨 150080)

摘 要: 身份-矢量(identity-vector, i-vector)方法作为说话人确认领域中的主流方法之一,能够通过学习总变化空间来获取有效的低维说话人特征——i-vector特征.但是当开发集数据不充足时,会导致学习到的总变化空间模型误差较大;同时,还无法有效确认此时的总变化空间是否因为预先设置的维度过高而学到了冗余信息.为此,本文将贝叶斯主成分分析(Bayesian Principal Component Analysis, BPCA)引入总变化空间的学习过程中,利用其来为总变化空间引入更多的先验信息,从而对开发集数据中包含的信息进行补充,并在先验信息的约束下削弱总变化空间中无效维的影响.实验结果表明,当开发集数据不充足时,相比于传统的总变化空间学习方法,BPCA方法能够有效提升说话人确认系统的识别性能.

关键词: 说话人确认; 身份-矢量(i-vector); 总变化空间; 贝叶斯主成分分析
中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112(2021)11-2186-09
电子学报URL: http://www.ejournal.org.cn **DOI:** 10.12263/DZXB.20200476

Bayesian Principal Component Analysis for I-Vector Speaker Verification

RONG Ya-feng[1], CHEN Chen[1,2], CHEN De-yun[1,2], HE Yong-jun[1]
(1. *School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, Heilongjiang* 150080, *China*;
2. *Postdoctoral Research Station of Computer Science and Technology, Harbin University of Science and Technology, Harbin, Heilongjiang* 150080, *China*)

Abstract: As one of the most important methods in speaker verification, the identity-vector (i-vector) approach can obtain effective low-dimensional i-vector by learning the total variability space (TVS). However, when there is no sufficient development data, it will lead to a large error in the learned TVS model. Meanwhile, it is difficult to determine whether there is redundancy in the learned TVS due to the high preset dimension. To solve the above problems, the Bayesian principal component analysis (BPCA) is introduced into the learning of the TVS. And this proposed method can introduce more prior information into the TVS to supply more information. Additionally, under the constraint of prior information, the influence of invalid dimension in the TVS can be weakened. The experimental results show that when the development data is insufficient, the BPCA method can effectively improve the performance compared with the traditional TVS learning methods.

Key words: speaker verification; i-vector; total variability space; Bayesian principal component analysis

1 引言

说话人确认是一项根据语音信息来确定某段语音是否来自特定说话人的身份鉴定技术,目前身份-矢量(identity-vector, i-vector)方法[1]是该领域中的主流方法之一.I-vector方法能够通过学习总变化空间,将说话人语音所对应的高维均值超矢量(mean supervector)[2]映射为低维i-vector特征.值得注意的是,对总变化空间的准确描述能够有效提升i-vector特征的区分能力,从而提升说话人确认系统的识别性能.因此,总变化空间学习作为i-vector方法中的关键任务之一,受到研究者们的高度重视.前端因子分析(Front-End Factor Analysis, FEFA)[1]作为最早被提出的总变化空间学习方法,

收稿日期:2020-05-19;修回日期:2020-11-09;责任编辑:孙瑶
基金项目:国家自然科学基金(No.62101163,No.61673142);黑龙江省自然科学基金(No.JJ2019JQ0013);黑龙江省博士后专项经费(No.LBH-Z20020);黑龙江省普通高校基本科研业务费专项资金(No.2020-KYYWF-0341);哈尔滨市杰出青年人才基金(No.2017RAYXJ013)

通过直接建立Baum-Welch统计量与i-vector特征之间的映射关系来学习总变化空间. 由于Baum-Welch统计量与均值超矢量具有等效性[3],此后相继出现了一系列直接对均值超矢量进行降维的方法. 例如主成分分析(Principal Component Analysis, PCA)[4]、因子分析(Factor Analysis, FA)[3]以及概率主成分分析(Probabilistic Principal Component Analysis, PPCA)[3,5]等. 以上方法更加关注对高维数据向低维空间进行映射时的残差处理. 此外,通过有监督的学习方式来充分利用说话人类别信息,能够使学习到的总变化空间更加具有区分能力. 例如,有监督概率主成分分析(Supervised Probabilistic Principal Component Analysis, SPPCA)方法[6]和偏最小二乘(Probabilistic Partial Least Squares, PLS)[7,8]等方法,能够使提取到的i-vector特征融合更多的类别信息. 此外,由于总变化空间中同时包含说话人信息与会话信息,为了去除会话信息的影响,在提取i-vector特征后还需要对其进行会话补偿[9]. 目前常用的会话补偿方法有线性判别分析(Linear Discriminant Analysis, LDA)[10]、类内协方差规整(Within-Class Covariance Normalization, WCCN)[11]以及扰动属性投影(Nuisance Attribute Projection, NAP)[12]等. 在识别阶段,则可以采用余弦距离打分(Cosine Distance Scoring, CDS)[1]方法或者概率线性判别分析(Probabilistic Linear Discriminant Analysis, PLDA)[13]方法对提取的i-vector特征进行识别.

随着深度学习在图像处理、语音识别等领域的快速发展,基于深度学习的方法正逐渐应用于说话人确认领域. 其中,d-vector方法[14]通过利用深度神经网络(Deep Neural Network, DNN)来提取帧级嵌入(embedding)特征,并将一段语音中全部帧级特征的均值作为这段语音的d-vector特征. X-vector方法[15,16]则利用时延神经网络(Time-Delay Neural Network, TDNN)[17]提取语音帧的上下文相关信息,然后采用统计池化层计算帧级特征的统计量,并从网络的最后一个隐藏层中提取出x-vector特征. 在此基础上,通过在帧级层上采用多尺度卷积方法[18],能够从不同的感受野中获取更多的说话人信息;通过将TDNN与统计池化层相结合[19],则能够获取更具表示能力的说话人特征. 此外,视觉几何组-中等(Visual Geometry Group-Medium, VGG-M)网络[20,21]与深度残差网络(Deep Residual Network, ResNet)[22,23]等方法均能够通过学习更复杂的网络架构来进行说话人特征表示.

深度神经网络方法需要依赖大量的开发集数据来学习网络参数,且随着网络层数的增加,模型学习的时间复杂度也会增大;因此当训练数据不充足时,利用复杂的深度网络结构很难学习到准确的说话人信息,这将严重影响系统的识别性能. 与此同时,目前的总变化空间学习方法虽然能够获得有效表示说话人信息的特征空间,但是对数据规模也有一定的需求,这在实际应用中存在着一定的问题. 考虑到说话人信息类别庞杂,用于模型训练的开发集数据不能涵盖全部的说话人信息,且每位说话人的数据也无法保证足够充足,因此,当开发集数据量不充足时,无法学习到区分能力足够强的总变化空间. 同时,由于总变化空间矩阵的维度需要凭借经验预先设定,因此无法确定学习到的模型是否因设置参数的维度过高而存在冗余信息,最终影响到说话人确认系统的识别性能.

基于以上分析,本文将贝叶斯主成分分析(Bayesian Principal Component Analysis, BPCA)[24,25]引入总变化空间的学习过程中,利用贝叶斯原理来为总变化空间引入更多的先验信息,以此来弥补开发集数据不充足的问题. 具体而言,需要对总变化空间矩阵定义一个高斯先验分布假设,从而在此先验假设的约束下,凸显有效维度的作用,并削弱无效维度的影响,进而增加总变化空间中包含的有效信息,最终提升其对说话人信息的表示能力.

2　基于PPCA的总变化空间学习方法

在说话人确认任务中,可以采用FEFA、FA和PPCA等方法来进行总变化空间学习. 值得注意的是,上述方法能够获得相近的识别性能[3],但FEFA方法的计算复杂度较高,而FA方法与PPCA方法的原理则较为类似,因此本节以PPCA方法为例,简要介绍其学习过程. PPCA方法通过直接建立均值超矢量与i-vector特征之间的映射关系来学习总变化空间,定义N段语音的均值超矢量为集合$\{\boldsymbol{M}_n \in \mathbb{R}^D;\ n=1,2,\cdots,N\}$,其中$N$为开发集语音数据总数,$D$为均值超矢量的维度. 则第$n$段语音对应的均值超矢量$\boldsymbol{M}_n$可以表示为以下线性关系:

$$\boldsymbol{M}_n = \boldsymbol{m} + \boldsymbol{T}\boldsymbol{w}_n + \boldsymbol{\varepsilon}_n \tag{1}$$

其中,$\boldsymbol{m} \in \mathbb{R}^D$为均值超矢量映射到i-vector特征的过程中产生的偏置;$\boldsymbol{w}_n \sim \mathcal{N}(0,\boldsymbol{I}) \in \mathbb{R}^R$为说话人隐变量,即i-vector特征,$R$为i-vector特征的维度;$\boldsymbol{T} \in \mathbb{R}^{D\times R}$为总变化空间矩阵;$\boldsymbol{\varepsilon}_n \sim \mathcal{N}\left(0,\sigma^2\boldsymbol{I}\right) \in \mathbb{R}^R$为残差矢量,$\sigma^2$为协方差矩阵的对角元素.

在PPCA方法中,总变化空间矩阵和协方差矩阵的估计方法与在FA中类似,可以采用期望最大化(Expectation Maximization, EM)算法对其进行参数更新. 不同之处在于,FA方法中的协方差矩阵为各向异性的矩阵,而PPCA中的协方差矩阵则为各向同性. 在进行PPCA方法的参数更新时,EM算法的求期望步骤(E步)需要估计出说话人隐变量$\boldsymbol{w}_n$的后验分布. 因此,需要计算在第n段语音对应的均值超矢量$\boldsymbol{M}_n$条件下,$\boldsymbol{w}_n$的后验

协方差矩阵 $\boldsymbol{L}$、后验均值 $\mathrm{E}\left[\boldsymbol{w}_n|\boldsymbol{M}_n\right]$ 与后验相关矩阵 $\mathrm{E}\left[\boldsymbol{w}_n\boldsymbol{w}_n^{\mathrm{T}}|\boldsymbol{M}_n\right]$，计算公式如下：

$$\boldsymbol{L}=\boldsymbol{I}+\frac{1}{\sigma^2}\boldsymbol{T}^{\mathrm{T}}\boldsymbol{T} \tag{2}$$

$$\mathrm{E}\left[\boldsymbol{w}_n|\boldsymbol{M}_n\right]=\frac{1}{\sigma^2}\boldsymbol{L}^{-1}\boldsymbol{T}^{\mathrm{T}}\left(\boldsymbol{M}_n-\boldsymbol{m}\right) \tag{3}$$

$$\mathrm{E}\left[\boldsymbol{w}_n\boldsymbol{w}_n^{\mathrm{T}}|\boldsymbol{M}_n\right]=\boldsymbol{L}^{-1}+\mathrm{E}\left[\boldsymbol{w}_n|\boldsymbol{M}_n\right]\mathrm{E}\left[\boldsymbol{w}_n^{\mathrm{T}}|\boldsymbol{M}_n\right] \tag{4}$$

在EM算法的最大化步骤（M步）中，则需要通过最大化 $\boldsymbol{w}_n$ 的对数似然函数，来更新总变化空间模型的参数. 基于此，参数 $\{\boldsymbol{T},\sigma^2\}$ 的更新公式可以分别表示为

$$\boldsymbol{T}=\left[\sum_{n=1}^{N}\left(\boldsymbol{M}_n-\boldsymbol{m}\right)\mathrm{E}\left[\boldsymbol{w}_n^{\mathrm{T}}|\boldsymbol{M}_n\right]\right]\left[\sum_{n=1}^{N}\mathrm{E}\left[\boldsymbol{w}_n\boldsymbol{w}_n^{\mathrm{T}}|\boldsymbol{M}_n\right]\right]^{-1} \tag{5}$$

$$\sigma^2=\frac{1}{DN}\sum_{n=1}^{N}\begin{bmatrix}\left(\boldsymbol{M}_n-\boldsymbol{m}\right)^{\mathrm{T}}\left(\boldsymbol{M}_n-\boldsymbol{m}\right)\\-\mathrm{E}\left[\boldsymbol{w}_n^{\mathrm{T}}|\boldsymbol{M}_n\right]\boldsymbol{T}^{\mathrm{T}}\left(\boldsymbol{M}_n-\boldsymbol{m}\right)\end{bmatrix} \tag{6}$$

经过E步与M步的反复迭代，上述参数最终会趋于收敛. 在识别阶段，对于任意一段语音的均值超矢量 $\boldsymbol{M}_n$，其对应的i-vector特征 $\boldsymbol{w}_n$ 可以表示为以下形式：

$$\boldsymbol{w}_n=\frac{1}{\sigma^2}\left(\boldsymbol{I}+\frac{1}{\sigma^2}\boldsymbol{T}^{\mathrm{T}}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^{\mathrm{T}}\left(\boldsymbol{M}_n-\boldsymbol{m}\right) \tag{7}$$

3 基于BPCA的总变化空间学习方法

针对开发集数据不充足而导致所学到的总变化空间模型误差过大，以及经验维度设置不合理而使模型中存在冗余信息的问题，本文将BPCA方法引入总变化空间的学习过程中，从而有效利用先验信息来弥补数据中缺少的信息量，且先验信息也能够对模型起到有效的约束作用.

3.1 总变化空间学习

与FA和PPCA方法一致，BPCA方法也将全部开发集数据的均值超矢量用作原始数据，第 n 段语音对应的均值超矢量表示方法与式（1）相同，且均值超矢量 $\boldsymbol{M}_n$ 在说话人隐变量 $\boldsymbol{w}_n$ 条件下服从高斯分布 $\boldsymbol{M}_n|\boldsymbol{w}_n\sim\mathcal{N}\left(\boldsymbol{m}+\boldsymbol{T}\boldsymbol{w}_n,\sigma^2\boldsymbol{I}\right)$. 在此基础上，为了利用先验信息来约束模型，本文所提方法对总变化空间矩阵 $\boldsymbol{T}=\left\{\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_r,\cdots,\boldsymbol{t}_R\right\}$ 的每一列定义一个独立的高斯先验分布，由精度超参数 α_r 控制，计算如下：

$$P(\boldsymbol{T}|\alpha)=\prod_{r=1}^{R}\left(\frac{\alpha_r}{2\pi}\right)^{\frac{D}{2}}\exp\left\{-\frac{1}{2}\alpha_r\boldsymbol{t}_r^{\mathrm{T}}\boldsymbol{t}_r\right\} \tag{8}$$

通过对精度超参数 α_r 的学习能够帮助确认总变化空间矩阵 $\boldsymbol{T}$ 中每个基 $\boldsymbol{t}_r$ 的有效性. 例如：当 α_r 具有一定数值时，能够保证总变化空间矩阵的第 r 列向量 $\boldsymbol{t}_r$ 有效，以此来凸显该维 $\boldsymbol{t}_r$ 对总变化空间的影响，使其学习到尽可能多的说话人信息；而当 α_r 趋于无穷大时，其对应的总变化空间矩阵的第 r 列向量 $\boldsymbol{t}_r$ 则趋近 $\boldsymbol{0}$，以此来抑制无效信息对模型的影响.

对于模型参数 $\{\boldsymbol{m},\boldsymbol{T},\alpha_r,\sigma^2\}$，也可以使用EM算法对其进行迭代求解. 在M步，需要最大化 $\boldsymbol{M}_n$ 的对数后验概率密度函数，它可表示为以下形式：

$$\begin{aligned}
&\sum_{n=1}^{N}\ln\left(P\left(\boldsymbol{M}_n;\,\boldsymbol{m},\boldsymbol{T},\sigma^2\right)P\left(\boldsymbol{T}|\alpha_r\right)\right)\\
&=\sum_{n=1}^{N}\ln\left(\sum_{\boldsymbol{w}_n}\left[P\left(\boldsymbol{M}_n,\boldsymbol{w}_n;\,\boldsymbol{m},\boldsymbol{T},\sigma^2\right)P\left(\boldsymbol{T}|\alpha_r\right)\right]\right)\\
&=\sum_{n=1}^{N}\ln\left(\sum_{\boldsymbol{w}_n}F\left(\boldsymbol{w}_n\right)\right)\frac{P\left(\boldsymbol{M}_n|\boldsymbol{w}_n;\,\boldsymbol{m},\boldsymbol{T},\sigma^2\right)P\left(\boldsymbol{w}_n\right)P\left(\boldsymbol{T}|\alpha_r\right)}{F\left(\boldsymbol{w}_n\right)}\\
&\geqslant\sum_{n=1}^{N}\sum_{\boldsymbol{w}_n}\ln\left(F\left(\boldsymbol{w}_n\right)\right)\frac{P\left(\boldsymbol{M}_n|\boldsymbol{w}_n;\,\boldsymbol{m},\boldsymbol{T},\sigma^2\right)P\left(\boldsymbol{w}_n\right)P\left(\boldsymbol{T}|\alpha_r\right)}{F\left(\boldsymbol{w}_n\right)}\\
&=\sum_{n=1}^{N}\mathrm{E}_{\boldsymbol{w}_n}\left[\ln\left(P\left(\boldsymbol{M}_n|\boldsymbol{w}_n;\,\boldsymbol{m},\boldsymbol{T},\sigma^2\right)\right)+\ln\left(P\left(\boldsymbol{T}|\alpha_r\right)\right)\right]\\
&\quad+\sum_{n=1}^{N}\mathrm{E}_{\boldsymbol{w}_n}\left[\ln\left(P\left(\boldsymbol{w}_n\right)\right)-\ln\left(F\left(\boldsymbol{w}_n\right)\right)\right]
\end{aligned} \tag{9}$$

其中，$F\left(\boldsymbol{w}_n\right)$ 为辅助函数，表示说话人隐变量 $\boldsymbol{w}_n$ 在 $\boldsymbol{M}_n$ 条件下的某种未知分布；$\mathrm{E}_{\boldsymbol{w}_n}[\boldsymbol{x}]$ 表示在 $\boldsymbol{w}_n$ 上对 $\boldsymbol{x}$ 求期望. 然后去掉式（9）中的参数无关项 $\ln P\left(\boldsymbol{w}_n\right)-\ln F\left(\boldsymbol{w}_n\right)$，目标函数简化后可以表示为

$$\begin{aligned}
&\mathcal{L}\left(\boldsymbol{M}_n|\boldsymbol{w}_n;\,\boldsymbol{m},\boldsymbol{T},\alpha_r,\sigma^2\right)\\
&=\sum_{n=1}^{N}\mathrm{E}_{\boldsymbol{w}_n}\left[\ln\left(P\left(\boldsymbol{M}_n|\boldsymbol{w}_n;\,\boldsymbol{m},\boldsymbol{T},\sigma^2\right)\right)+\ln\left(P\left(\boldsymbol{T}|\alpha_r\right)\right)\right]\\
&=\sum_{n=1}^{N}\mathrm{E}_{\boldsymbol{w}_n}\left[-\frac{1}{2}\ln\left|\sigma^2\right|\right]+\sum_{n=1}^{N}\sum_{r=1}^{R}\left[\frac{D}{2}\ln\alpha_r-\frac{1}{2}\alpha_r\boldsymbol{t}_r^{\mathrm{T}}\boldsymbol{t}_r-c_2\right]\\
&\quad+\sum_{n=1}^{N}\mathrm{E}_{\boldsymbol{w}_n}\left[-\frac{1}{2}\left(\boldsymbol{M}_n-\boldsymbol{m}-\boldsymbol{T}\boldsymbol{w}_n\right)^{\mathrm{T}}\sigma^{-2}\left(\boldsymbol{M}_n-\boldsymbol{m}-\boldsymbol{T}\boldsymbol{w}_n\right)-c_1\right]
\end{aligned} \tag{10}$$

其中，c_1、c_2 为常数. 分别对式（10）中的参数 $\{\boldsymbol{m},\boldsymbol{T},\alpha_r,\sigma^2\}$ 求偏导数并令其为 $\boldsymbol{0}$，即可求得每个参数的更新公式. 对于偏置 $\boldsymbol{m}$，其偏导数可以表示为

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{m}}=-\sum_{n=1}^{N}\frac{1}{\sigma^2}\left(\boldsymbol{m}-\boldsymbol{M}_n\right)+\sum_{n=1}^{N}\frac{1}{\sigma^2}\boldsymbol{T}\mathrm{E}\left[\boldsymbol{w}_n\right]=\boldsymbol{0} \tag{11}$$

对式（11）进行求解，可以得到 $\boldsymbol{m}$ 的更新公式如下：

$$\boldsymbol{m}=\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{M}_n \tag{12}$$

其中，$\boldsymbol{m}$ 为全部开发集数据对应的均值超矢量的均值.

对于精度超参数 α_r，其偏导数可以表示为

$$\frac{\partial\mathcal{L}}{\partial\alpha_r}=\frac{D}{2\alpha_r}-\frac{1}{2}\boldsymbol{t}_r^{\mathrm{T}}\boldsymbol{t}_r=0 \tag{13}$$

对式(13)进行求解,可以得到α_r的更新公式如下:

$$\alpha_r = \frac{D}{\boldsymbol{t}_r^{\mathrm{T}}\boldsymbol{t}_r} \tag{14}$$

对于总变化空间矩阵$\boldsymbol{T}$,其偏导数可以表示为

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \boldsymbol{T}} &= \sum_{n=1}^{N}\frac{1}{\sigma^2}\left(\boldsymbol{M}_n-\boldsymbol{m}\right)\mathrm{E}\left[\boldsymbol{w}_n^{\mathrm{T}}|\boldsymbol{M}_n\right] \\ &\quad -\sum_{n=1}^{N}\frac{1}{\sigma^2}\boldsymbol{T}\mathrm{E}\left[\boldsymbol{w}_n\boldsymbol{w}_n^{\mathrm{T}}|\boldsymbol{M}_n\right]-\boldsymbol{T}\boldsymbol{A} \\ &= \boldsymbol{0}\end{aligned} \tag{15}$$

其中,$\boldsymbol{A}=\mathrm{diag}\left(\alpha_1,\alpha_2,\cdots,\alpha_R\right)$; $r=1,2,\cdots,R$; $\mathrm{diag}(\cdot)$表示以$\cdot$为对角元素的对角矩阵. 对式(15)进行求解,可以得到$\boldsymbol{T}$的更新公式如下:

$$\boldsymbol{T}=\left[\sum_{n=1}^{N}\left(\boldsymbol{M}_n-\boldsymbol{m}\right)\mathrm{E}\left[\boldsymbol{w}_n^{\mathrm{T}}|\boldsymbol{M}_n\right]\right]\left[\sum_{n=1}^{T}\mathrm{E}\left[\boldsymbol{w}_n\boldsymbol{w}_n^{\mathrm{T}}|\boldsymbol{M}_n\right]+\sigma^2\boldsymbol{A}\right]^{-1} \tag{16}$$

对比式(5)与式(16)可以发现,式(16)中增加了一项$\sigma^2\boldsymbol{A}$,通过矩阵$\boldsymbol{A}$中的先验信息来约束总变化空间矩阵$\boldsymbol{T}$,从而增强总变化空间中有效信息的表示并抑制无效信息对其的影响,保证总变化空间模型的有效性.

对于协方差σ^2,其倒数的偏导数可以表示为

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \sigma^{-2}} &= \frac{1}{2}\sum_{n=1}^{N}\left[\sigma^2\boldsymbol{I}-\left\|\boldsymbol{M}_n-\boldsymbol{m}\right\|^2\right] \\ &\quad +\sum_{n=1}^{N}\left(\boldsymbol{M}_n-\boldsymbol{m}\right)\mathrm{E}\left[\boldsymbol{w}_n^{\mathrm{T}}|\boldsymbol{M}_n\right]\boldsymbol{T}^{\mathrm{T}} \\ &\quad -\frac{1}{2}\sum_{n=1}^{N}\boldsymbol{T}\mathrm{E}\left[\boldsymbol{w}_n\boldsymbol{w}_n^{\mathrm{T}}|\boldsymbol{M}_n\right]\boldsymbol{T}^{\mathrm{T}} \\ &= 0\end{aligned} \tag{17}$$

对式(17)进行求解,可以得到σ^2的更新公式如下:

$$\begin{aligned}\sigma^2 = \frac{1}{DN}\sum_{n=1}^{N}\Big\{&-2\mathrm{E}\left[\boldsymbol{w}_n^{\mathrm{T}}|\boldsymbol{M}_n\right]\boldsymbol{T}^{\mathrm{T}}\left(\boldsymbol{M}_n-\boldsymbol{m}\right) \\ &+\left\|\boldsymbol{M}_n-\boldsymbol{m}\right\|^2+\mathrm{tr}\left[\mathrm{E}\left[\boldsymbol{w}_n\boldsymbol{w}_n^{\mathrm{T}}|\boldsymbol{M}_n\right]\boldsymbol{T}^{\mathrm{T}}\boldsymbol{T}\right]\Big\}\end{aligned} \tag{18}$$

其中,$\mathrm{tr}(\cdot)$为求迹运算.

上文介绍了EM算法中M步求解参数$\{\boldsymbol{m},\boldsymbol{T},\alpha_r,\sigma^2\}$的过程. 下面将给出EM算法中E步的求解过程,即求取说话人隐变量$\boldsymbol{w}_n$的后验均值$\mathrm{E}\left[\boldsymbol{w}_n|\boldsymbol{M}_n\right]$与后验相关矩阵$\mathrm{E}\left[\boldsymbol{w}_n\boldsymbol{w}_n^{\mathrm{T}}|\boldsymbol{M}_n\right]$的过程. 每段语音的i-vector特征在均值超矢量$\boldsymbol{M}_n$条件下的后验分布可以表示为

$$\begin{aligned}&P\left(\boldsymbol{w}_n|\boldsymbol{M}_n\right) \\ &\propto P\left(\boldsymbol{M}_n|\boldsymbol{w}_n\right)P\left(\boldsymbol{w}_n\right) \\ &\propto \exp\left\{\frac{1}{\sigma^2}\boldsymbol{w}_n^{\mathrm{T}}\boldsymbol{T}^{\mathrm{T}}-\frac{1}{2}\boldsymbol{w}_n^{\mathrm{T}}\left[\boldsymbol{I}+\frac{1}{\sigma^2}\boldsymbol{T}^{\mathrm{T}}\boldsymbol{T}\right]\boldsymbol{w}_n\right\}\end{aligned} \tag{19}$$

对于任意的高斯分布$\mathcal{N}\left(\boldsymbol{x};\boldsymbol{\mu}_x,\boldsymbol{\Sigma}_x\right)$,均可以写为以下形式:

$$\begin{aligned}&P\left(\boldsymbol{x};\boldsymbol{\mu}_x,\boldsymbol{\Sigma}_x\right) \\ &\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}_x\right)^{\mathrm{T}}\boldsymbol{\Sigma}_x^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}_x\right)\right\} \\ &\propto \exp\left\{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\mu}_x-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Sigma}_x^{-1}\boldsymbol{x}\right\}\end{aligned} \tag{20}$$

对比式(19)与式(20),记$\boldsymbol{L}=\boldsymbol{I}+\sigma^{-2}\boldsymbol{T}^{\mathrm{T}}\boldsymbol{T}$,则说话人隐变量$\boldsymbol{w}_n$的后验均值$\mathrm{E}\left[\boldsymbol{w}_n|\boldsymbol{M}_n\right]$与后验相关矩阵$\mathrm{E}\left[\boldsymbol{w}_n\boldsymbol{w}_n^{\mathrm{T}}|\boldsymbol{M}_n\right]$可以表示为

$$\mathrm{E}\left[\boldsymbol{w}_n|\boldsymbol{M}_n\right]=\frac{1}{\sigma^2}\boldsymbol{L}^{-1}\boldsymbol{T}^{\mathrm{T}}\left(\boldsymbol{M}_n-\boldsymbol{m}\right) \tag{21}$$

$$\mathrm{E}\left[\boldsymbol{w}_n\boldsymbol{w}_n^{\mathrm{T}}|\boldsymbol{M}_n\right]=\boldsymbol{L}^{-1}+\mathrm{E}\left[\boldsymbol{w}_n|\boldsymbol{M}_n\right]\mathrm{E}\left[\boldsymbol{w}_n^{\mathrm{T}}|\boldsymbol{M}_n\right] \tag{22}$$

根据以上推导内容,下面给出基于BPCA的总变化空间模型学习算法,如算法1所示. 参数可以经过E步和M步的反复迭代,最终趋于收敛.

算法1　基于BPCA的总变化空间模型学习算法

输入:$\boldsymbol{M}_n$为均值超矢量;R为i-vector维度;I为迭代次数

输出:参数$\boldsymbol{m},\boldsymbol{T},\sigma^2$

1. 随机初始化$\boldsymbol{T},\sigma^2$
2. 用式(12)计算$\boldsymbol{m}$
3. for i=1 to I do
4. 　用式(21)与式(22)更新后验均值$\mathrm{E}\left[\boldsymbol{w}_n|\boldsymbol{M}_n\right]$
5. 　与后验相关矩阵$\mathrm{E}\left[\boldsymbol{w}_n\boldsymbol{w}_n^{\mathrm{T}}|\boldsymbol{M}_n\right]$
6. 　用式(14)、式(16)与式(18)更新参数α_r、$\boldsymbol{T}$、σ^2
7. end

3.2　I-vector特征提取

在学习总变化空间模型之后,需要从每一段语音中提取出表征说话人身份的i-vector特征. 利用均值超矢量$\boldsymbol{M}$和总变化空间矩阵$\boldsymbol{T}$来提取i-vector特征,于是每个说话人的i-vector特征,即说话人隐变量$\boldsymbol{w}_n$的后验均值,可以表示为

$$\boldsymbol{w}_n=\frac{1}{\sigma^2}\left(\boldsymbol{I}+\frac{1}{\sigma^2}\boldsymbol{T}^{\mathrm{T}}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^{\mathrm{T}}\left(\boldsymbol{M}_n-\boldsymbol{m}\right) \tag{23}$$

从式(23)可以发现,i-vector特征通过更新后的参数$\boldsymbol{T}$将先验信息融入其中. 因此,由式(23)提取到的i-vector特征中也包含了总变化空间的先验信息,使得有效信息得以凸显且无效信息被抑制,以此来增加i-vector特征的表示能力,从而提升说话人识别系统的性能.

4　实验与分析

4.1　数据库与前端实验设置

本文实验采用TIMIT数据库[26],其为包含630位说话人(438男,192女)的英文语音库,该数据库被广泛应

用于说话人识别的研究中[27,28]. 每位说话人包含10段语音,共计6300段语音. 由于官方并未对该数据库划分开发集与评估集数据,本文将针对说话人确认任务对其进行划分. 将前530位说话人的5300段语音用作开发集数据,后100位共计1000段语音用作评估集数据. 将评估集中每位说话人的9段语音用作注册语音,1段语音用作测试语音. 实验共计10000次测试,其中包含100次目标测试与9900次非目标测试.

在进行前端特征提取之前,需要先对各说话人语音进行语音活动检测(Voice Activity Detection, VAD)处理,去除语音中的静音部分,然后再进行特征提取. 前端特征采用梅尔频率倒谱系数(Mel-Frequency Cepstral Coefficients, MFCC)特征,先对所有语音数据进行预加重和分帧处理,预加重系数为0.95,帧长为25ms,帧移为10ms. 对所有语音数据提取13维MFCC基本特征,然后再对其进行一阶、二阶差分计算,并拼接得到39维声学特征. 在模型训练阶段,高斯混合模型的混合分量为512个.

4.2 性能对比与分析

本节将对比本文提出的BPCA方法与其他两种方法(FA、PPCA)的识别性能. 采用等错误率(Equal Error Rate, EER)与最小检测代价函数(Minimum Detection Cost Function, Min DCF)来衡量系统的性能,其中Min DCF参数设置为$C_{\text{miss}}=1$,$C_{\text{fa}}=1$,$P_{\text{target}}=0.01$. FA、PPCA与BPCA方法的总变化空间维度均设置为400维[1],并采用LDA对i-vector特征进行会话补偿,LDA的维度设置为200维. 在识别阶段,采用CDS与PLDA两种方法进行说话人确认匹配,当采用PLDA作为后端分类器时,需要先对提取到的i-vector特征进行长度规整(Length Normalization, LN)[29],然后训练PLDA模型用于分类,PLDA分类器子空间维度设置为200维. 根据上述实验设置,不同方法的实验结果如表1所示.

表1 不同方法的性能对比

方法	分类器	EER/%	Min DCF
FA	CDS	11.70	0.73
FA+LDA		7.00	0.63
PPCA		6.00	0.74
PPCA+LDA		5.00	0.64
BPCA		5.19	0.68
BPCA+LDA		3.39	0.55
FA	PLDA	3.97	0.59
FA+LDA		2.78	0.52
PPCA		3.04	0.66
PPCA+LDA		3.00	0.60
BPCA		3.00	0.60
BPCA+LDA		2.47	0.57

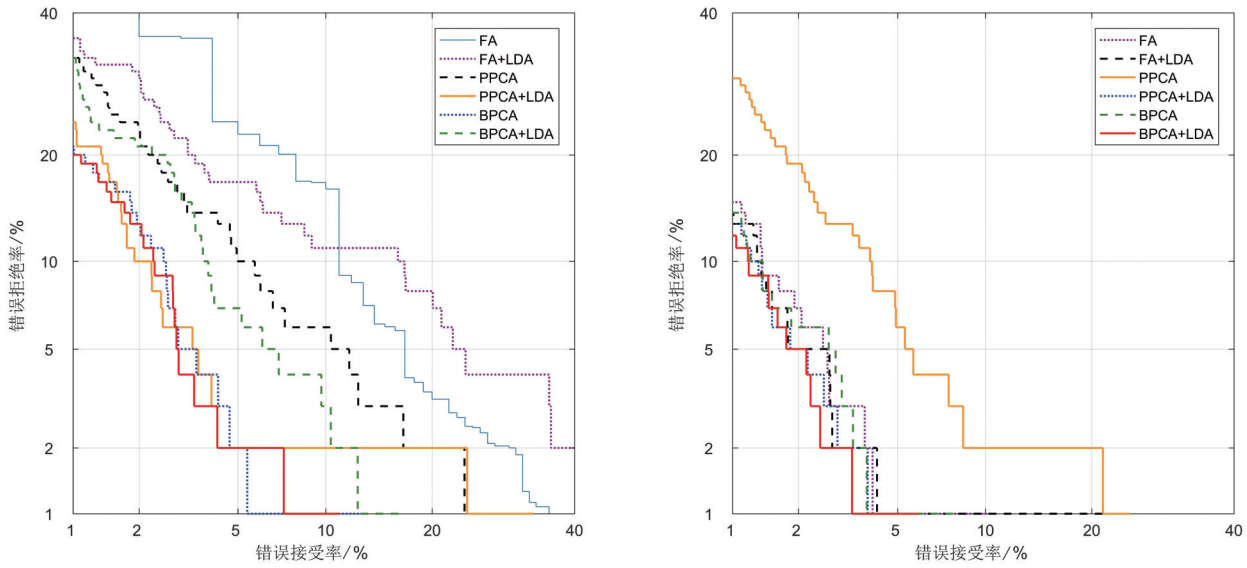
从表1可以看出:①当采用CDS进行分类时,与FA、PPCA方法相比,BPCA方法的EER相对降低了55.64%和13.5%,其性能明显优于FA和PPCA方法,原因在于,先验假设能够帮助学习更有效的总变化空间矩阵,即凸显总变化空间矩阵的某些有效维度,并抑制某些无效维度;②当采用PLDA作为分类器时,与FA、PPCA方法相比,BPCA方法的EER相对降低了24.42%和1.33%,由于PLDA作为概率型分类器在学习说话人子空间时也对说话人隐变量做出了先验假设,能够对i-vector特征中的会话信息进行进一步的补偿,从而提升了系统的识别性能;③除了采用PLDA分类器进行会话补偿外,当额外对i-vector特征进行会话补偿处理后,系统可以获得更好的识别性能,特别是BPCA+LDA+PLDA方法使系统性能达到最优. 由以上实验结果可以验证,BPCA方法能够使系统取得更优的识别性能,此收益正是来自于引入的先验分布假设,其能使总变化空间在学习过程中充分利用各个维度的信息,从而使得总变化空间矩阵的有效维度得以凸显且无效维度被抑制,进而提升识别性能.

不同方法的检测错误权衡(Detection Error Tradeoff, DET)曲线如图1所示. 从图1中可以看出:无论使用CDS还是PLDA作为后端分类器,BPCA+LDA方法对应的DET曲线均在最下方. 由此可见,BPCA方法在说话人确认系统中能够取得更优的性能,进一步验证了本文所提的总变化空间学习方法的优势.

4.3 Hinton图对比与分析

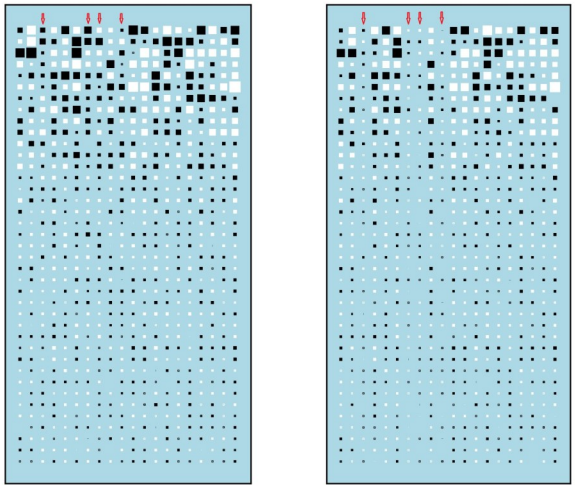
考虑到BPCA方法是在PPCA方法的基础上,通过引入精度超参数α_r来控制总变化空间矩阵$\boldsymbol{T}$中各列数据的有效性,因此为了从更直观的角度来分析BPCA方法的性能,本节将对比经PPCA与BPCA方法训练后所得矩阵$\boldsymbol{T}$的Hinton图. Hinton图能够将矩阵中的每个元素表示为一个正方形,其中白色表示正值,黑色表示负值,并且正方形的面积与其对应元素的大小成正比. 由于矩阵$\boldsymbol{T}=\left[\boldsymbol{T}_1,\cdots,\boldsymbol{T}_r,\cdots,\boldsymbol{T}_{512}\right]^{\mathrm{T}}$的维度过高不利于数据的可视化,且每一个$\boldsymbol{T}_c\in\mathbb{R}^{39\times 400}$均与GMM的第$c$个高斯分量对应,因此只需选取其中一个高斯分量对应的数据$\boldsymbol{T}_c$即可. 同时,考虑到无法全部展示具有400列的$\boldsymbol{T}_c$,从其400列数据中随机选取20列,最终得到维度为39×20的矩阵$\tilde{\boldsymbol{T}}$. 根据上述设置,PPCA方法与BPCA方法所对应的矩阵$\tilde{\boldsymbol{T}}$的Hinton图如图2所示.

对比图2的(*a*)与(*b*)可以发现:图2(*b*)中第3、7、8、10列(红色箭头指示)的正方形明显小于图2(*a*)中对应的正方形,即相比于PPCA方法,BPCA方法所得矩阵$\tilde{\boldsymbol{T}}$中对应的这几列数值很小. 这是因为当超参数α_r趋于无穷大时,其对应的总变化空间矩阵的第r列向量趋近于$\boldsymbol{0}$,从而通过α_r数值的增加来削减其所对应维(图



(*a*)采用CDS分类方法的DET曲线　　(*b*)采用PLDA分类方法的DET曲线

图1　不同方法的DET曲线对比



(*a*) PPCA方法　　(*b*) BPCA方法

图2　PPCA与BPCA方法的部分总变化空间矩阵对应的Hinton图

中第3、7、8、10列)的作用,进而抑制这些列来减少其对i-vector特征提取的影响.

4.4　开发集数据规模对性能影响的对比与分析

本节将验证不同规模的开发集数据对系统性能的影响. 实验将开发集数据分为四种规模,随机抽取原始开发集数据的1/K(K=1,2,4,8)作为新的开发集数据. 为了描述方便,分别将全部数据、1/2开发集数据、1/4开发集数据和1/8开发集数据简记为数据集Ⅰ~Ⅳ. 随着开发集数据规模的缩小,其中包含的说话人类别数目减少,总变化空间不需要设置过高的维度. 因此,实验中需要采用不同的维度设置,具体维度设置如表2所示.

表2　不同规模开发集数据的空间维度设置

数据集	数据规模(类)	总变化空间维度	LDA维度	PLDA子空间维度
数据集Ⅰ	530	400	200	200
数据集Ⅱ	265	200	120	120
数据集Ⅲ	132	70	60	50
数据集Ⅳ	66	60	50	30

性能评价指标仍采用等错误率(EER)与最小检测代价函数(Min DCF). 根据以上设置,使用不同的开发集数据来学习总变化空间模型,并在4.1节中介绍的评估集上进行验证,实验结果如表3所示.

从表3可以得出如下结论.

(1)从整体上来看,随着数据规模的减小,不同方法的性能均有所下降,但BPCA方法仍能保持相对更优的性能,这是因为BPCA方法在总变化空间上引入高斯先验假设,弥补开发集数据不足的问题,使得在数据规模较小时仍能获得良好的识别性能.

(2)当采用CDS作为分类方法时,无论是否进行会话补偿,在不同数据规模下BPCA方法性能均优于其他方法,说明了先验信息对总变化空间模型学习的有效性.

(3)当采用PLDA作为分类器时,BPCA方法仍能保持相对更优的性能,这再次验证了先验信息对总变化空间模型的学习起到了有效作用. 同时,相比于使用CDS作为识别方法,各方法的性能均有所提升,这是因为PLDA方法在学习说话人子空间时也引入了先验分布假设,使得i-vector特征更加具有区分能力.

由以上实验结果可以验证,当开发集数据不充足

表3 不同规模开发集数据下不同方法的性能对比

方法	分类器	数据集Ⅰ		数据集Ⅱ		数据集Ⅲ		数据集Ⅳ	
		EER/%	Min DCF	EER/%	Min DCF	EER/%	Min DCF	EER/%	Min DCF
FA	CDS	11.70	0.73	10.23	0.75	17.41	0.98	19.00	0.98
FA+LDA		7.00	0.63	7.69	0.74	16.00	0.96	16.38	0.97
PPCA		6.00	0.74	7.34	0.80	12.00	0.93	13.77	0.96
PPCA+LDA		5.00	0.64	6.63	0.77	11.49	0.91	13.47	0.96
BPCA		5.19	0.68	6.36	0.76	11.39	0.91	13.47	0.96
BPCA+LDA		3.39	0.55	5.00	0.67	11.32	0.90	12.71	0.92
FA	PLDA	3.97	0.59	5.51	0.75	12.20	0.92	15.75	0.97
FA+LDA		2.78	0.52	3.82	0.74	12.17	0.92	14.00	0.97
PPCA		3.04	0.66	3.90	0.65	12.00	0.97	13.41	0.96
PPCA+LDA		3.00	0.60	3.42	0.62	11.00	0.93	12.79	0.95
BPCA		3.00	0.60	3.01	0.69	11.18	0.94	12.37	0.93
BPCA+LDA		2.47	0.57	3.00	0.68	10.00	0.90	11.79	0.92

时，相比于传统的总变化空间学习方法，基于BPCA方法的说话人确认系统能够取得更优的性能.

此外，经观察发现，表3中数据集Ⅱ到Ⅲ所对应的EER明显升高. 针对这一现象，通过如下实验分析其原因. 在数据集Ⅱ到Ⅲ之间以较小幅度缩减数据规模，将其划分为230类、170类与150类，它们所对应的总变化空间维度分别设置为200、150与120，LDA与PLDA的维度分别设置为120、120与80. 仍然采用LDA与PLDA的原因是：根据上述实验得知使用LDA+PLDA方法进行会话补偿时系统性能相对较好. 根据以上设置，使用不同的开发集数据集来学习总变化空间模型，在4.1节中介绍的评估集上进行验证，FA、PPCA与BPCA方法的性能变化如图3所示.

从图中可以看出，随着开发集数据规模的减小，说话人识别系统的性能评价指标EER逐渐上升，尤其是

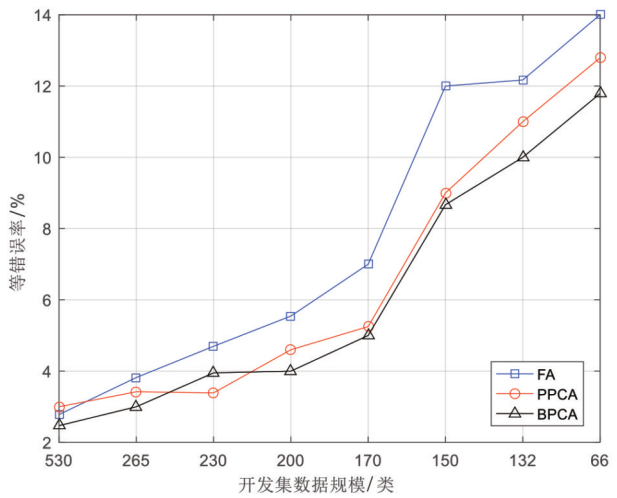


图3 不同规模开发集数据下不同方法的性能变化

在数据规模减小至170类之后，两种方法的EER明显上升. 由此可见，当数据集的规模减小至一定量时，说话人识别系统的性能会发生质的改变，导致EER的增加较明显，即性能降低较明显.

5 结论

本文提出了一种基于BPCA的总变化空间学习方法，此方法能够有效缓解由开发集数据不充足导致的总变化空间模型误差较大的问题，同时，其也能够根据先验信息的约束而有效补偿总变化空间矩阵中存在的冗余信息. 具体而言，本文所提方法通过对总变化空间矩阵定义高斯先验分布假设，来约束各维度数据对模型的影响，从而既能使总变化空间学习到尽可能多的说话人信息，又可以抑制无效信息对其的影响，进而提升模型对说话人特征的表示能力，保证系统有较好的识别性能. 实验表明，无论开发集数据充足与否，基于BPCA方法的说话人确认系统性能均能够优于其他传统的总变化空间学习方法.

参考文献

[1] Dehak N, Kenny P J, Dehak R, et al. Front-end factor analysis for speaker verification[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2011, 19(4): 788 – 798.

[2] Campbell W M, Sturim D E, Reynolds D A. Support vector machines using GMM supervectors for speaker verification[J]. IEEE Signal Processing Letters, 2006, 13(5): 308 – 311.

[3] Vestman V, Kinnunen T. Supervector compression strategies to speed up i-vector system development[A]. The

Speaker and Language Recognition Workshop (Odyssey 2018) [C] Les Sables d'Olonne, France: ISCA, 2018. 357 – 364.

[4] Lei Z C, Yang Y C. Maximum likelihood i-vector space using PCA for speaker verification[A]. Proceedings of Twelfth Annual Conference of the International Speech Communication Association[C]. Florence, Italy: ISCA, 2011. 2725 – 2728.

[5] Tipping M E, Bishop C M. Probabilistic principal component analysis[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 1999, 61(3): 611 – 622.

[6] Lei Y, Hansen J H L. Speaker recognition using supervised probabilistic principal component analysis[A]. Proceedings of Eleventh Annual Conference of the International Speech Communication Association[C]. Makuhari, Japan: ISCA, 2010. 382 – 385.

[7] Chen C, Han J Q, Pan Y L. Speaker verification via estimating total variability space using probabilistic partial least squares[A]. Proceedings of Eighteenth Annual Conference of the International Speech Communication Association[C]. Stockholm, Sweden: ISCA, 2017. 1537 – 1541.

[8] Chen C, Han J Q. Partial least squares based total variability space modeling for I-vector speaker verification[J]. Chinese Journal of Electronics, 2018, 27(6): 1229 – 1233.

[9] 韩纪庆, 张磊, 郑铁然. 语音信号处理(第3版)[M]. 北京: 清华大学出版社, 2019.
Han J Q, Zhang L, Zheng T R. Speech Signal Processing (3rd ed) [M]. Beijing, China: Tsinghua University Press, 2019. (in Chinese)

[10] Fisher R A. The use of multiple measurements in taxonomic problems[J]. Annals of Eugenics, 1936, 7(2): 179 – 188.

[11] Hatch A O, Kajarekar S, Stolcke A. Within-class covariance normalization for SVM-based speaker recognition [A]. Proceedings of Ninth International Conference on Spoken Language Processing[C]. Pittsburgh, Pennsylvania,USA: ISCA, 2006. 1471 – 1474.

[12] Campbell W M, Sturim D E, Reynolds D A, et al. SVM based speaker verification using a GMM supervector kernel and NAP variability compensation[A]. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing[C]. Toulouse, France: IEEE, 2006. 97 – 100.

[13] Prince S J D, Elder J H. Probabilistic linear discriminant analysis for inferences about identity [A]. Proceedings of IEEE International Conference on Computer Vision[C]. Rio de Janeiro, Brazil: IEEE, 2007. 1 – 8.

[14] Variani E, Lei X, McDermott E, et al. Deep neural networks for small footprint text-dependent speaker verification[A]. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing[C]. Florence, Italy: IEEE, 2014. 4052 – 4056.

[15] Nyder D, Garcia-Romero D, Povey D, et al. Deep neural network embeddings for text-independent speaker verification[A]. Proceedings of Eighteenth Annual Conference of the International Speech Communication Association [C]. Stockholm, Sweden: ISCA, 2017. 999 – 1003.

[16] Snyder D, Garcia-Romero D, Sell G, et al. X-vectors: robust DNN embeddings for speaker recognition[A]. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing[C]. Calgary, AB, Canada: IEEE, 2018. 5329 – 5333.

[17] LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural Computation, 1989, 1(4): 541 – 551.

[18] Gu B, Guo W, Dai L R, et al. An improved deep neural network for modeling speaker characteristics at different temporal scales[A]. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing [C]. Barcelona, Spain: IEEE, 2020. 6814 – 6818.

[19] Hong Q B, Wu C H, Wang H M, et al. Statistics pooling time delay neural network based on x-vector for speaker verification[A]. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing[C]. Barcelona, Spain: IEEE, 2020. 6849 – 6853.

[20] Chatfield K, Simonyan K, Vedaldi A, et al. Return of the devil in the details: Delving deep into convolutional nets [A]. Proceedings of the British Machine Vision Conference[C]. Nottingham, UK: BMVA, 2014. 1 – 12.

[21] Nagrani A, Chung J S, Zisserman A. VoxCeleb: A large-scale speaker identification dataset[A]. Proceedings of Eighteenth Annual Conference of the International Speech Communication Association[C]. Stockholm, Sweden: ISCA, 2017. 2616 – 2620.

[22] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition[C]. LasVegas, NV, USA: IEEE, 2016. 770 – 778.

[23] Chung J S, Nagrani A, Zisserman A. Voxceleb2: deep speaker recognition[A]. Proceedings of Nineteenth Annual Conference of the International Speech Communication Association[C]. Hyderabad, India: ISCA, 2018.

1086 – 1090.
[24] Bishop C M. Bayesian PCA[J]. Advances in Neural Information Processing Systems, 1999, 11(2): 382 – 388.
[25] Bishop C M. Machine Learning and Pattern Recognition [M]. New York, USA: Springer, 2006.
[26] Jankowski C, Kalyanswamy A, Basson S, et al. NTIMIT: A phonetically balanced, continuous speech, telephone bandwidth speech database[A]. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing[C]. Albuquerque, NM, USA: IEEE, 1990. 109 – 122.
[27] 蒋晔, 唐振民. 短语音说话人辨认的研究[J]. 电子学报, 2011, 39(4): 953 – 957.
Jiang Y, Tang Z M. Research on the speaker identification based on short utterance[J]. Acta Electronica Sinica, 2011, 39(4): 953 – 957. (in Chinese)
[28] 张二华, 王明合, 唐振民. 加性噪声条件下鲁棒说话人确认[J]. 电子学报, 2019, 47(6): 1244 – 1250.
Zhang E H, Wang M H, Tang Z M. Robust speaker verification under additive noise condition[J]. Acta Electronica Sinica, 2019, 47(6): 1244 – 1250. (in Chinese)
[29] Garcia-Romero D, Espy-Wilson C Y. Analysis of i-vector length normalization in speaker recognition systems[A]. Proceedings of Twelfth Annual Conference of the International Speech Communication Association[C]. Florence, Italy: IEEE, 2011. 249 – 252.

作者简介

彤娅峰 女,1997年出生.哈尔滨理工大学计算机科学与技术学院硕士研究生.主要研究方向为说话人识别、语音信号处理等.
E-mail:rongyafeng908@163.com

陈 晨 女,1990年出生.哈尔滨理工大学计算机科学与技术学院讲师、博士后、硕士生导师.主要研究方向为语音信号处理、音频信息分析、说话人识别等.
E-mail:chenc@hrbust.edu.cn

陈德运(通讯作者) 男,1962年出生.哈尔滨理工大学计算机科学与技术学院教授、博士生导师.主要研究方向为模式识别、机器学习等.
E-mail:chendeyun@hrbust.edu.cn

何勇军 男,1980年出生.哈尔滨理工大学计算机科学与技术学院教授、博士生导师.主要研究方向为语音信号处理、图像处理等.
E-mail:holywit@163.com