文章编号：1673-5196(2021)05-0093-06

基于多特征 I-Vector 的说话人识别算法

赵 宏*，岳鲁鹏，常兆斌，王伟杰

（兰州理工大学 计算机与通信学院，甘肃 兰州 730050）

摘要：针对单一声学特征无法精准高效地辨识说话人身份的问题，提出了一种基于多特征 I-Vector 的说话人识别算法.该算法首先采集不同的声学特征并将其构成一个高维特征向量，然后通过主成分分析法有效地剔除高维特征向量的关联，确保各种特征之间正交化，最后采用概率线性判别分析进行建模和打分，并在一定程度上降低空间维度.在 TIMIT 语料库上利用 Kaldi 进行实验，算法运行结果表明，该算法较当前流行的基于 I-Vector 的单一梅尔频率倒谱系数和感知线性预测系数的特征系统在等错误率上分别提高了 8.18%和 1.71%，在模型训练时间上分别减少了 60.4%和 47.5%，具有更好的识别效果和效率.

关键词：说话人识别算法；多特征 I-Vector；主成分分析；概率线性判别分析；Kaldi

中图分类号：TN391 **文献标志码**：A

Speaker recognition algorithm based on multi-featured I-Vector

ZHAO Hong，YUE Lu-peng，CHANG Zhao-bin，WANG Wei-jie

(College of Computer and Communication，Lanzhou Univ. of Tech.，Lanzhou 730050，China)

Abstract：Aiming at the problem of inaccurate and inefficient speaker recognition presented by single acoustic feature，a speaker recognition algorithm was proposed based on multi-featured I-Vector. Firstly，different acoustic feature vectors were extracted and combined into a high-dimensional feature vector.Then principal components analysis (PCA) was used to effectively remove the correlation of these feature vectors，so that the features became orthogonalized. Finally，probabilistic linear discriminant analysis(PLDA) was used for modeling and scoring，which led to reduce the spatial dimension to a certain degree.Experiments were carried out on TIMIT corpus in combination with Kaldi speech recognition toolkit，and the results compared with the single-featured systems including Mel-frequency cepstral coefficients (MFCC) and perceptual linear predictive (PLP) coefficients based on I-Vector，the equal error rate (EER) of the purposed algorithm were increased by 8.18%and 1.71%，respectively；the model training time were decreased respectively by 60.4% and 47.5%，respectively.Therefore，the purposed algorithm has betterspeaker recognition performance and efficiency.

Key words：speaker recognition algorithm；multi-featured I-Vector；principal components analysis；probabilistic linear discriminant analysis；Kaldi speech recognition toolkit

说话人识别(speaker recognition，SR)又称声纹识别(voiceprint recognition，VR)，是一种根据语音信息对讲话者身份进行辨识的生物特征识别技术.该技术的作用原理是根据语音中能够描述和反映讲话者独特特征的信息，通过信息识别等一系列先进成熟的技术高效、智能地辨识讲话者身份[1-2].与面部识别等其他识别技术相比，说话人识别优势突出，它不仅操作便捷、成本较低，并且识别精度高，已经在金融、军事等多个行业得到了大力推广和积极应用，应用前景非常可观.

目前，关于说话人识别的研究，学者们一般从前端处理和后端处理这两点出发展开探讨和分析.

在前端处理上，Esun 等[3]从语音特征入手，通过对声学特征参数的分析，提出了一种基于 DBN-

收稿日期：2019-12-19
基金项目：国家自然科学基金(51668043)，赛尔网络下一代互联网技术创新项目(NGII20160311，NGII20160112)
作者简介：赵 宏(1971-)，男，甘肃西和人，博士，教授，博导.
Email：594286500@qq.com

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

GMM 混合模型的深层高斯相关超矢量(deep Gaussian correlation super vector,DGCS)特征,从而降低了系统的错误率和训练复杂度.Agrawal 等[4]从时-频特征入手,通过对语音频谱图进行分析,提出了使用语音频谱图学习初始滤波器,而使用残差谱图学习后续滤波器的方法,从而抑制了易受噪声干扰的区域,提高了系统的鲁棒性.李作强等[5]从相位特征入手,通过对相位信息的分析,设计出了相位特征参数提取算法,不仅使得系统识别率进一步提升,同时也明显增强了系统鲁棒性.张二华等[6]从噪声干扰入手,通过对噪声环境下的说话人确认系统分析,探索出了一种以分区约束非负矩阵分解为基础的语音去噪策略,克服了在噪声比较大的场合下无法精准辨识讲话者的不足,取得了较好的效果.

在后端处理上,Reynolds 等[7]提出了高斯混合模型-统一背景模型(Gaussian mixture model-universal background model,GMM-UBM)框架,尝试将此框架应用于说话人验证系统中,取得了良好的成果.Han 等[8]在进行广泛深入的研究之后探索出了联合因子分析(joint factor analysis,JFA)技术,将讲话者、信道之间所存在的不同固定于 GMM 超矢量高维空间的两个子空间内,从而使得所有高斯成分之间均存在一定关联,由此实现信息的传输与共享.Dehak 等[9]通过深入研究提出了一个低维、长度固定的身份矢量(identity vector,I-Vector),使得无论是何种长度的语音段,均能够通过长度固定的 I-Vector 进行描述[10],由此令 I-Vector 的训练复杂度进一步降低,因子向量的推理过程更简单.由于 I-Vector 系统不仅存储了说话人信息,也含有信道信息,所以一般需要包括线性判别分析(linear discriminant analysis,LDA)[11]、干扰属性映射(nuisance attribute projection,NAP)[12]、类内协方差归一化(within class covariance normalization,WCCN)[13]和概率线性判别分析(probabilistic linear discriminant analysis,PLDA)[14]等后端区分技术为信道提供一定补偿.由于 PLDA 的补偿能力更好,故使用 PLDA 为 I-Vector 提供信道补偿,从而减少信道信息对说话人识别任务的影响,进一步提高 I-Vector 系统对说话人的区分能力.

以上说话人识别的处理方法各有所长,相比而言,前端处理法的优势在于可以较好地体现信号中的时序特性,但对于语音数据的质量和数量要求较高,鲁棒性较差.后端处理法的优势在于具有良好的灵活性,其不足主要表现为两个方面:一方面运算量过大,另一方面是出现过拟合现象的可能性比较大.

针对上述问题,本文提出一种结合前端和后端的联合优化算法.在前端处理上,选择梅尔频率倒谱系数(Mel-frequency cepstral coefficient,MFCC)和感知线性预测系数(perceptual linear predictive,PLP)两种声学特征组合,然后采用 PCA 对新的动态特征参数降维;在后端处理上,将降维后的新的动态特征参数放入 I-Vector 说话人识别系统中,并结合 PLDA 对说话人和信道变化进行建模和打分,从而克服多种声学特征简单组合的局限性,不仅有助于提高系统识别率,还明显改善了系统鲁棒性.

1 基于多特征 I-Vector 的说话人识别模型

基于多特征 I-Vector 说话人识别模型如图 1 所示,分为数据预处理、特征提取、特征处理、模型训练和打分判决五个步骤.其中,数据预处理利用 Kaldi[15]读取语音信号并创建相关文件;特征提取是将语音信号 MFCC 和 PLP 两种特征分别提取并进行融合;特征处理利用 PCA 来解决特征提取阶段的冗余问题;模型训练分为 UBM 模型训练和训练特征两个部分;打分判决采用 PLDA 对变换处理后的测试特征向量以及说话人的注册特征向量予以决策评分.

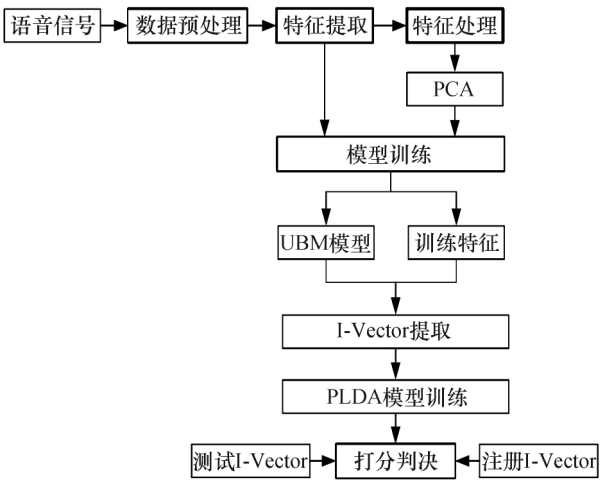

图 1 算法框架图

Fig.1 Algorithm framework

1.1 数据预处理

数据预处理包括语音语料的准备和数据部分的准备.关于语音语料的准备,本文采用 TIMIT 语音数据库来进行实验;数据部分的准备利用 Kaldi 生成 spk2utt、utt2spk 和 wav.scp 等文件.

1.2 特征提取

说话人识别本质是一种模式识别,因此合理地

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

选择提取的特征显得尤为重要[16].本文将 MFCC 和 PLP 融合作为混合特征参数,计算公式为

$$F_{m\times n}=[(A_1,A_2,\cdots,A_m),(B_1,B_2,\cdots,B_n)] \tag{1}$$

式中:A 和 B 分别表示 MFCC 和 PLP 特征参数;m 和 n 分别表示 MFCC、PLP 的特征参数阶数.

1.3　特征处理

由于不同特征之间不仅存在冗余信息,而且包含系统贡献率较低的维度,为此本文使用 PCA 来保持总特征空间的基向量不相关和正交,换言之,将有效剔除特征向量间的内在关系,同时要去除各种冗余信息.

假设输入数据矩阵为 $\boldsymbol{\chi}_i$ 且平均值为 0,在单位向量 $\boldsymbol{\omega}$ 方向上的协方差矩阵计算如下式所示:

$$\sigma_\omega=\frac{1}{n}\sum_{i=1}^{n}|\boldsymbol{\chi}_i\times\boldsymbol{\omega}|^2=\frac{1}{n}\boldsymbol{\omega}^{\mathrm{T}}\left(\sum_{i=1}^{n}\boldsymbol{\chi}_i\boldsymbol{\chi}_i^{\mathrm{T}}\right)\boldsymbol{\omega} \tag{2}$$

式中:$\sum_{i=1}^{n}\boldsymbol{\chi}_i\boldsymbol{\chi}_i^{\mathrm{T}}=\boldsymbol{\chi}\boldsymbol{\chi}^{\mathrm{T}}$ 表示 x 的协方差矩阵.同时采用拉格朗日乘子法计算 σ_ω 的最大值,计算公式为

$$\max\{\boldsymbol{\omega}^{\mathrm{T}}\Sigma\boldsymbol{\omega}\};\quad \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}=1 \tag{3}$$

$$f(\boldsymbol{\omega})=\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}+\lambda(1-\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}) \tag{4}$$

$$\frac{\partial f}{\partial\boldsymbol{\omega}}=2\Sigma\boldsymbol{\omega}-2\lambda\boldsymbol{\omega}=0\rightarrow\Sigma\boldsymbol{\omega}=\lambda\boldsymbol{\omega} \tag{5}$$

在式(5)中,让其成立的条件是投影矩阵的协方差实现极大值.因此,式(5)等同于求解特征值 λ 以及特征向量 $\boldsymbol{\omega}$.关于 σ_ω,计算公式为

$$\sigma_\omega=\lambda\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}=\lambda \tag{6}$$

在式(6)中,λ 随着 σ_ω 增大而增大.因此前 M 个最大特征值对应特征向量的投影矩阵 $\boldsymbol{\omega}_{\mathrm{PCA}}$ 的计算公式为

$$\boldsymbol{\omega}_{\mathrm{PCA}}=[\boldsymbol{\omega}_1\quad\boldsymbol{\omega}_2\quad\cdots\quad\boldsymbol{\omega}_M] \tag{7}$$

由式(1)直接拼接起来的混合参数过于冗余,此外,在语音中低维特征矩阵容易受噪声因素的干扰,从而导致声纹识别的正确率降低[17],因此本文在特征提取后,采用 PCA 计算投影矩阵,特征矩阵 $\boldsymbol{F}$ 的协方差矩阵计算公式为

$$\boldsymbol{\sigma}_{\boldsymbol{F}}=\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F} \tag{8}$$

在式(8)中,特征矩阵的协方差 $\boldsymbol{\sigma}_{\boldsymbol{F}}$ 是对角矩阵,计算公式为

$$\boldsymbol{\sigma}_{\boldsymbol{F}}=\begin{bmatrix}\boldsymbol{\iota}_1\\\boldsymbol{\iota}_2\\\vdots\\\boldsymbol{\iota}_t\end{bmatrix}\begin{bmatrix}\lambda_1&0&\cdots&0\\0&\lambda_2&\cdots&0\\\vdots&\vdots&&\vdots\\0&0&\cdots&\lambda_t\end{bmatrix}[\boldsymbol{\iota}_1\quad\boldsymbol{\iota}_2\quad\cdots\quad\boldsymbol{\iota}_t] \tag{9}$$

式中:$\boldsymbol{\iota}$ 是特征值 λ 对应的归一化特征向量.假设 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_t$,那么投影矩阵 $\boldsymbol{\iota}_{\mathrm{PCA}}$ 可以由前 p 个最大特征值对应的特征向量所生成,计算公式为

$$\boldsymbol{\iota}_{\mathrm{PCA}}=[\boldsymbol{\iota}_1\quad\boldsymbol{\iota}_2\quad\cdots\quad\boldsymbol{\iota}_p] \tag{10}$$

在式(10)中,$\boldsymbol{\iota}_{\mathrm{PCA}}$ 用于将所有特征向量映射到不相关的特征空间,将提取的特征经过 PCA 变换后用于模型训练,计算公式为

$$\dot{\boldsymbol{F}}=\boldsymbol{\iota}_{\mathrm{PCA}}^{\mathrm{T}}\boldsymbol{F} \tag{11}$$

1.4　模型训练

将上述准备好的语音语料和文件作为模型的输入端,并根据 UBM 模型和训练特征,从而生成 I-Vector 说话人模型.Dhanush 等[18]认为在 JFA 技术的信道子空间中,与说话人相关的信息中也同时含有一些干扰信息,如信道信息等,而这同时也是造成分割子空间建模缺乏精准性的一项重要原因.鉴于此,本文决定采用精准度高且便于计算的 I-Vector 模型进行建模,其计算公式为

$$M=\boldsymbol{\mu}+\boldsymbol{T}w \tag{12}$$

式中:$\boldsymbol{\mu}$ 表示与说话者无关且信道无关的超向量,可通过 UBM 的均值超向量进行拼接组成;$\boldsymbol{T}$ 是一个低秩矩阵;w 则为服从标准正态分布的随机向量.声学特征 $\boldsymbol{\chi}_t^{(i)}$ 的计算公式为

$$\boldsymbol{\chi}_t^{(i)}\sim\sum_k\gamma_{kt}^{(i)}N(\mu_k+\boldsymbol{T}\boldsymbol{X}^{(i)},\Sigma_k) \tag{13}$$

式中:$k(k=1,2,\cdots,K)$ 为高斯分量总数;$\boldsymbol{X}^{(i)}$ 为第 i 个说话人的语音段所对应的说话人身份向量;μ_k 和 Σ_k 分别为第 k 个高斯分量均值和协方差;$\gamma_{kt}^{(i)}=p(k\mid x_t^{(i)})$ 的计算公式为

$$\gamma_{k\mathrm{t}}^{(i)}=\frac{\widetilde{\omega}_kN_{\mathrm{UBM}}(\boldsymbol{\chi}_t^{(i)};\mu_k,\Sigma_k)}{\sum_{k=1}^{K}\widetilde{\omega}_kN_{\mathrm{UBM}}(\chi_t^{(i)};\mu_k,\Sigma_k)} \tag{14}$$

式中:$\widetilde{\omega}_k$ 为第 k 个高斯分量的权重.

根据一段语音,利用 UBM 便捷高效地推导出统计量,其具体计算公式为

$$N_k^{(i)}=\sum_t\gamma_{k,\mathrm{t}}^{(i)} \tag{15}$$

$$F_k^{(i)}=\sum_t\gamma_{k,t}^{(i)}\boldsymbol{\chi}_t^{(i)} \tag{16}$$

$$S_k^{(i)}=\sum_t\gamma_{k,t}^{(i)}\boldsymbol{\chi}_t^{(i)}\boldsymbol{\chi}_t^{(i)\ \mathrm{T}} \tag{17}$$

式中:$N_k^{(i)}$、$F_k^{(i)}$ 和 $S_k^{(i)}$ 分别表示第 i 个说话人的语音段在第 k 个高斯分量上的零阶、一阶和二阶统计量.

1.5　打分判决

PLDA 是对 I-Vector 进一步的因子分析[19].Garcia-Romero 等[20]在传统 PLDA 模型的基础上提出了一种更为简单的 PLDA 模型,并且证明了这

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

种模型的区分性比传统 PLDA 模型更好.计算公式为

$$\boldsymbol{P}_i = m + \boldsymbol{\Phi}\beta + \varepsilon_\gamma \tag{18}$$

式中:m 为训练数据全局均值;$\boldsymbol{\Phi}$ 为描述说话人类间差异的子空间矩阵;β 是满足标准正态分布特征的隐含因子,且与说话人身份相关;ε_γ 是对角矩阵 $\boldsymbol{\Sigma}$ 的残余项.PLDA 模型参数 $\{m,\boldsymbol{\Phi},\beta\}$ 的值可通过文献[21]中的 EM 算法求得.

在决策打分环节,依次用 η_1、η_2 进行描述,两者的对数似然比的计算公式为

$$\text{score} = \lg \frac{p(\eta_1,\eta_2 \mid R_s)}{p(\eta_1 \mid R_d)\,p(\eta_2 \mid R_d)} \tag{19}$$

式中:R_s 表示 η_1 和 η_2 来自同一个说话人;R_d 表示 η_1 和 η_2 来自不同的说话人.

2 实验结果与分析

2.1 实验环境

本文选用 Kaldi 进行实验,实验环境如表 1 所列.

表 1 实验环境

Tab.1 Experimental environment

参数	数值
CPU	2.6 GHz Intel Core i7
内存	16 GB
操作系统	MacOS Mojave 10.14
测试平台	Kaldi-trunk

2.2 实验数据集

为对算法性能的可靠性、精准性进行严格全面的检验,本文在 TIMIT 语音数据库的 630 人中筛选出了 462 人组成一个专门的训练集,余下的 168 人则组成验证集和测试集.在训练集中,每人有 10 句话,而这意味着需要对 4 620 条语音进行规范合理的训练.而在测试集中,每人有 8 句话,以此完成验证集的训练;还有 2 句话,用于测试说话人.通过 Kaldi 预处理并准确提取语音信号的特性,选用的特征包含 20 维 MFCC(包括第 0 级系数)的特征矢量和 12 维 PLP 的特征矢量.

2.3 评价标准

为评估说话人识别系统的性能,通过等错误率(equal error rate,EER)这一指标[22]来描述和反映系统性能.它主要反映了错误接受率(FAR)与错误拒绝率(FRR)相等时的错误率.一般来讲,等错误率的值愈小,系统性能愈佳.计算公式为

$$\text{FAR} = \frac{N_{\text{FA}}}{N_{\text{IRA}}} \tag{20}$$

$$\text{FRR} = \frac{N_{\text{FR}}}{N_{\text{GRA}}} \tag{21}$$

$$\text{EER} = \text{FAR} = \text{FRR} \tag{22}$$

式中:N_{FA} 表示匹配失败却判为匹配成功的说话人数量;N_{IRA} 表示匹配失败的总数量;N_{FR} 表示匹配成功却判为匹配失败的说话人数量;N_{GRA} 表示匹配成功的总数量.

2.4 参数选择

特征提取时,首先进行参数调优,目前针对网络模型参数的选择只能在先验知识的基础上进行多次实验,通过对比不同参数实验的结果,确定模型的最终参数.在参数确定后,本文采集到的特征进行全面细致的语音活动检测,同时采用了声纹特征训练高斯混合模型.各相关参数如表 2 所列.

表 2 实验参数

Tab.2 Experimental parameters

参数	数值
采样率/kHz	16
语音分帧	帧长 25 ms,帧移 10 ms
预加重系数	0.97
窗函数	汉明窗
Mel 滤波器个数	40

在 GMM-UBM 模型中,设置 1 024 个高斯混合数,全局差异空间矩阵在参与计算时可直接进行随机初始化处理,反复重复迭代 7 次,一直持续到获得符合要求的全局差异空间矩阵.

2.5 性能比较

首先,通过 Kaldi 在 TIMIT 数据库上对两种特征进行直接拼接成 33 维的特征向量;然后,通过主成分分析法(principal components analysis,PCA)作进一步降维处理,由此确认出区分度最高的特征.不同降维程度对 I-Vector 基线系统的性能影响如图 2 所示.

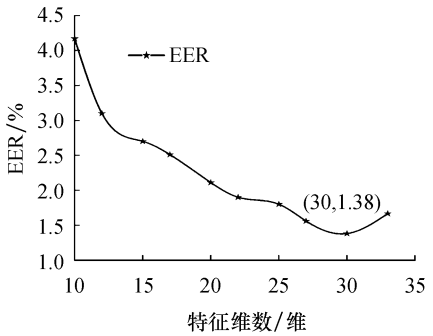

图 2 不同维度对基线系统性能的影响

Fig.2 Impact of different dimensions on baseline system performance

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

分析图 2 可知,当特征维数在 30 维之前时,I-Vector 基线系统的 EER 一直呈现下降趋势,系统性能不断提高,且当特征维数在第 30 维左右时,I-Vector 基线系统的 EER 达到最低.当特征维数超过 30 维后,I-Vector 基线系统的 EER 呈上升趋势,使得系统性能不断降低.因此,本文研究中经过处理后的动态特征维数为前 30 维.

为验证不同特征对 I-Vector 基线系统下的性能影响,本文分别选择 MFCC、PLP、直接相连和本文提出的多特征变换算法进行对比实验,不同性能对比实验结果见如表 3.

表 3　不同特征性能比较

Tab.3　Performance comparison of different characteristics

特征种类	维数	EER/%
MFCC	20	8.33
PLP	13	2.50
直接相连	33	3.47
本文算法	30	0.833

由表 3 可以看出,本文提出算法的 EER 要明显低于直接相连、单一 MFCC 特征和 PLP 特征算法.这说明对于 I-Vector 说话人识别系统,本文所提算法相较于单一 MFCC 和 PLP 特征系统的 EER 分别提高了 8.18%和 1.71%.

为了全面准确地了解和掌握本系统的说话人识别性能,本文设计并实施了多次试验,关于全部说话人的识别方案如表 4 所列,关于考虑性别差异的识别方案可参考表 5 和表 6.

表 4　系统不考虑区分性别的差异

Tab.4　Gender differences are not considered in the system

实验设置	识别方案	EER/%	时间/h
MFCC	性别无关	8.33	1.06
PLP	性别无关	2.50	0.80
本文算法	性别无关	0.833	0.42

表 5　系统识别男声的差异

Tab.5　Differences in system recognition of male voice

实验设置	识别方案	EER/%	时间/h
MFCC	男性	11.67	0.46
PLP	男性	2.77	0.38
本文算法	男性	1.38	0.21

表 6　系统识别女声的差异

Tab.6　Differences in system recognition of female voice

实验设置	识别方案	EER/%	时间/h
MFCC	女性	9.72	0.72
PLP	女性	3.50	0.50
本文算法	女性	0.694	0.27

可以发现在系统不考虑区分性别差异的情况下,对于 I-Vector 说话人识别系统,本研究所提算法相较于单一 MFCC 和 PLP 特征系统的 EER 分别提高了 8.18%和 1.71%;在模型训练时间上分别减少了 60.4%和 47.5%.只考虑系统识别男声的情况下,对于 I-Vector 说话人识别系统,本文所提算法相较于单一 MFCC 和 PLP 特征系统的 EER 分别提高了 11.6%和 1.43%;在模型训练时间上分别减少了 54.3%和 44.7%.只考虑系统识别女声的情况下,对于 I-Vector 说话人识别系统,本文所提算法相较于单一 MFCC 和 PLP 特征系统的 EER 分别提高了 10.0%和 2.91%;在模型训练时间上分别减少了 62.5%和 46.0%.以上实验数据能够充分证明本文提出算法的严谨有效性.

3　结语

本文提出一种基于多特征 I-Vector 的说话人识别算法.首先采集 MFCC 和 PLP 等特征并将其进行有效融合,解决了单一声学特征无法精准高效地辨识说话人身份的问题;然后对融合的动态特征采用 PCA 进行变换;最后将变换后的特征送入 I-Vector/PLDA 基线系统进行建模和打分判决.通过在 TIMT 语料库上进行实验,算法运行结果表明了本文算法的可行性和高效性,但在复杂噪声下识别性能有一定的降低.

在今后的研究中,将考虑在语料中加入一定量的噪声数据对其进行训练,并采用不同的噪声分离算法,从而降低噪声对说话人识别效果的影响.

参考文献:

[1] 仲伟峰,方　祥,范存航,等.深浅层特征及模型融合的说话人识别[J].声学学报,2018,43(2):263-272.

[2] 艾斯卡尔·肉孜,王　东,李蓝天,等.说话人识别中的分数域语速归一化[J].清华大学学报(自然科学版),2018,58(4):337-341.

[3] SUN E L,GU T,XIE K,*et al*.Text-independent speaker identification based on deep Gaussian correlation supervector [J]. International Journal of Speech Technology,2019,22(2):449-457.

[4] AGRAWAL P, GANAPATHY S. Unsupervised modulation filter learning for noise-robust speech recognition [J]. The Journal of the Acoustical Society of America,2017,142(3):1686-1692.

[5] 李作强,高　勇.基于 CFCC 和相位信息的鲁棒性说话人辨识[J].计算机工程与应用,2015,51(17):228-232.

[6] 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.

[7] REYNOLDS D A,QUATIERI T F,DUNN R B.Speaker veri-

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

fication using adapted Gaussian mixture models [J]. Digital Signal Processing, 2000, 10(1/2/3): 19-41.

[8] HAN L, LI T, ZHENG W, *et al*. Recognition of voiceprint using deep neural network combined with support vector machine [C]//International Conference on Intelligent and Interactive Systems and Applications. Bangkok Thailand: Springer, 2019: 3-11.

[9] 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, 2010, 19(4): 788-798.

[10] 栗志意, 张卫强, 何 亮, 等. 基于总体变化子空间自适应的 i-vector 说话人识别系统研究 [J]. 自动化学报, 2014, 40(8): 1836-1840.

[11] KANAGASUNDARAM A, DEAN D, VOGT R, *et al*. Weighted LDA techniques for i-vector based speaker verification [C]//IEEE International Conference on Acoustics, Speech and Signal Processing. Kyoto, Japan: IEEE, 2012: 4781-4784.

[12] LUO D, XIA L, ZHANG C, *et al*. Automatic pronunciation evaluation in high-states English speaking tests based on deep neural network models [C]//International Conference on Artificial Intelligence and Big Data. Chengdu: IEEE, 2019: 124-128.

[13] TANG H, LEI Z, HUANG Z, *et al*. The GMM and I-Vector systems based on spoofing algorithms for speaker spoofing detection [C]//Chinese Conference on Biometric Recognition. Zhuzhou: Springer, 2019: 502-510.

[14] MACHLICA L, ZAJIC Z. An efficient implementation of probabilistic linear discriminant analysis [C]//IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver, Canada: IEEE, 2013: 7678-7682.

[15] POVEY D, GHOSHAL A, BOULIANNE G, *et al*. The Kaldi speech recognition toolkit [C]//IEEE 2011 workshop on automatic speech recognition and understanding. Hilton Waikoloa Village, Big Island, Hawaii, US: IEEE, 2011: 1-4.

[16] 吴文昭. 基于多约简 α-GMM 和 SVM 的说话人确认 [J]. 兰州理工大学学报, 2015, 41(4): 113-116.

[17] 周 萍, 沈 昊, 郑凯鹏. 基于 MFCC 与 GFCC 混合特征参数的说话人识别 [J]. 应用科学学报, 2019, 37(1): 24-32.

[18] DHANUSH B K, SUPARNA S, AARTHY R, *et al*. Factor analysis methods for joint speaker verification and spoof detection [C]//IEEE International Conference on Acoustics, Speech and Signal Processing. New Orleans, LA, USA: IEEE, 2017: 5385-5389.

[19] 王 昕, 张洪冉. 基于 DNN 处理的鲁棒性 I-Vector 说话人识别算法 [J]. 计算机工程与应用, 2018, 54(22): 167-172.

[20] ARCIA-ROMERO D, ESPY-WILSON C Y. Analysis of i-vector length normalization in speaker recognition systems [C]//Twelfth Annual Conference of the International Speech Communication Association. Florence, Italy: [s.n.], 2011: 249-252.

[21] PRINCE S J D, ELDER J H. Probabilistic linear discriminant analysis for inferences about identity [C]//IEEE 11th International Conference on Computer Vision. Florence, Italy: IEEE, 2007: 1-8.

[22] NIST speaker recognition evaluation [EB/OL]. [2019-09-05]. http://www.itl.nist.gov/iad/mig/tests/spk/2010/index.html.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net