基于判别邻域嵌入算法的说话人识别

梁春燕*① 袁文浩① 李艳玲② 夏 斌① 孙文珠①
①(山东理工大学计算机科学与技术学院 淄博 255049)
②(内蒙古师范大学计算机与信息工程学院 呼和浩特 010022)

摘 要：该文提出一种基于判别邻域嵌入(DNE)算法的说话人识别。判别邻域嵌入算法作为流形学习方法的一种，可以通过构建邻接图获取数据的局部邻域结构信息；同时该算法可以充分利用类间判别信息，具有更强的判别能力。在美国国家标准技术研究院2010年说话人识别评测(NIST SRE 2010)电话-电话核心测试集上的实验结果表明了该算法的有效性。

关键词：说话人识别；总变化因子分析；邻域保持嵌入；判别邻域嵌入
中图分类号：TP391.42 **文献标识码**：A **文章编号**：1009-5896(2019)07-1774-05
DOI: 10.11999/JEIT180761

Speaker Recognition Using Discriminant Neighborhood Embedding

LIANG Chunyan① YUAN Wenhao① LI Yanling② XIA Bin① SUN Wenzhu①
①(*College of Computer Science and Technology, Shandong University of Technology, Zibo* 255049, *China*)
②(*College of Computer and Information Engineering, Inner Mongolia Normal University, Hohhot* 010022, *China*)

Abstract: Discriminant Neighborhood Embedding (DNE) algorithm is introduced into the speaker recognition system. DNE is a manifold learning approach and aims at preserving the local neighborhood structure on the data manifold. As well, DNE has much more power in discrimination by sufficiently using the between-class discriminant information. The experimental results on the telephone-telephone core condition of the NIST 2010 Speaker Recognition Evaluation (SRE) dataset indicate the effectiveness of DNE algorithm.

Key words: Speaker recognition; Total variability factor analysis; Neighborhood Preserving Embedding (NPE); Discriminant Neighborhood Embedding (DNE)

1 引言

说话人识别技术是利用语音信号中所包含的说话人特征信息，识别某个说话人身份的技术[1–3]。近年来，因子分析(Factor Analysis, FA)[4–8]成为说话人识别领域的重要技术，其中，总变化因子分析在说话人识别中占据主导地位。总变化因子分析在空间建模时，将说话人信息和信道信息看作一个整体空间，即总变化空间。通过对包含说话人信息和信道信息的高斯混合模型(Gaussian Mixed Model, GMM)高维均值超向量在总变化空间上进行投影，得到低维的因子向量i-vector，并在其基础上进行分类建模和判决[6,9]。

但是，总变化因子分析技术存在一定的不足。总变化因子分析技术实质是概率主成分分析(Probabilistic Principal Component Analysis, PPCA)[10]的一种，只能够反映数据的整体结构。为了克服上述缺点，邻域保持嵌入(Neighborhood Preserving Embedding, NPE)等流形学习算法成功应用于说话人识别[11–14]，有效地保持了数据流形上局部邻域结构信息，从而提高了说话人识别性能。然而，NPE算法[11,15]在进行数据线性重构时，仅保持了同类(同一说话人)样本的局部结构，未考虑不同类(不同说话人)样本之间的判别性，而判别信息

收稿日期：2018-08-03；改回日期：2019-01-21；网络出版：2019-02-24
*通信作者： 梁春燕 liangchunyan@sdut.edu.cn
基金项目：国家自然科学基金(11704229, 61701286, 61562068)，山东省自然科学基金(ZR2017LA011, ZR2015FL003, ZR2017MF047)，山东省高等学校科技计划项目(J17KA078)，内蒙古自然科学基金项目(2015MS0629)
Foundation Items: The National Natural Science Foundation of China (11704229, 61701286, 61562068), The Shandong Provincial Natural Science Foundation (ZR2017LA011, ZR2015FL003, ZR2017MF047), The Project of Shandong Province Higher Educational Science and Technology Program (J17KA078), The Natural Science Foundation of Inner Mongolia Autonomous Region of China (2015MS0629)

对说话人识别是非常重要的。

针对以上问题，本文提出一种基于判别邻域嵌入(Discriminant Neighborhood Embedding, DNE)算法的说话人识别系统。判别邻域嵌入算法通过结合邻域和类的信息，不仅能保持类内(同一说话人)样本数据的局部邻域结构，同时强调类间(不同说话人)样本数据间的判别信息，使得不同类样本的嵌入向量相互分类，因而具有更强的判别能力。

本文安排如下：第2节简要介绍基于i-vector的邻域保持嵌入算法，第3节是本文提出的判别邻域嵌入算法，第4节是在NIST SRE 2010电话-电话核心测试集上的实验结果，第5节是结论。

2　基于i-vector的邻域保持嵌入技术

2.1　总变化因子分析

给定一段语音数据，基于总变化空间的说话人和信道相关的GMM均值超向量可表示为

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

其中，$\boldsymbol{m}$是与说话人和信道无关的超向量，一般用通用背景模型(Universal Background Model, UBM)的均值超向量来表征；$\boldsymbol{T}$是一个低阶的总变化矩阵，维数为$CF\times R_T$，其中C是GMM高斯数，F为特征维数，R_T是总变化矩阵中所包含的特征向量的个数；$\boldsymbol{w}$是一个随机向量，服从标准正态分布$N(0,\ I)$。向量$\boldsymbol{w}$中的元素为总变化因子(Total Variability Factors, TVF)，称向量$\boldsymbol{w}$为identity vector，简称i-vector。对于训练$\boldsymbol{T}$和提取i-vector的具体过程详见参考文献[6]。

2.2　邻域保持嵌入(NPE)

给定训练数据集$\boldsymbol{X}=\{\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ \cdots,\ \boldsymbol{x}_N\}$，其中$\boldsymbol{x}_i\in\mathrm{R}^L$, $i=1, 2, \cdots, N$。$\boldsymbol{x}_i$是i-vector向量，且具备说话人身份的标注信息。NPE的目的是在降维的同时，保持数据集固有的局部邻域流形结构不变。它寻找一个最优的映射变换矩阵$\boldsymbol{P}=[\boldsymbol{p}_1,\ \boldsymbol{p}_2,\ \cdots,\ \boldsymbol{p}_K]$，将$\mathrm{R}^L$空间的数据嵌入映射到一个相对低维的向量空间$\mathrm{R}^K(K<L)$中。数据点$\boldsymbol{x}_i$在$\mathrm{R}^K$空间中表示为邻域保持嵌入向量$\boldsymbol{y}_i$，且$\boldsymbol{y}_i=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{x}_i$。

对于邻域保持嵌入空间矩阵$\boldsymbol{P}$的训练过程详见文献[11,15]。

将邻域保持嵌入向量使用支持向量机(Support Vector Machine, SVM)分类建模，并选用余弦核作为核函数。在进行SVM建模时，需要进行信道补偿。常用的信道补偿技术有2种：类内协方差规整(Within-Class Covariance Normalization, WCCN)[16]和线性判别分析(Linear Discriminant Analysis, LDA)[17]，这两个方法以不同的准则对信道进行补偿，同时使用能够起到更优的效果。

将NPE算法作用在i-vector上，可以同时保持数据样本的整体结构和局部邻域结构，因此极大地提高了系统性能[11]。但邻域保持嵌入算法在进行线性重构时，没有对邻域作类别判断和区别对待，使得不同类的邻近数据点在投影时存在相互重叠的情况，从而影响了识别性能。

3　判别邻域嵌入(DNE)

判别邻域嵌入(DNE)是一种有效的子空间学习方法，已成功应用于人脸识别[18,19]。通过与总变化因子分析相结合，能够从整体和局部更全面地反映数据的结构，同时能强调类的判别信息，对邻域作类别判断和区别对待。将DNE算法应用于说话人识别的思想如下：

给定训练数据集$\boldsymbol{X}=\{\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ \cdots,\ \boldsymbol{x}_N\}$，其中$\boldsymbol{x}_i\in\mathrm{R}^L$, $i=1, 2, \cdots, N$。$\boldsymbol{x}_i$是i-vector向量，且给定说话人身份的标注信息。DNE在降维的同时，能保持类内数据样本固有的局部邻域流形结构不变，并最大限度地使类间数据样本彼此远离，有效提高降维后的分类效果。它寻找一个最优的映射变换矩阵$\boldsymbol{A}=[\boldsymbol{a}_1,\ \boldsymbol{a}_2, \cdots, \boldsymbol{a}_K]$，将$\mathrm{R}^L$空间的数据嵌入映射到一个相对低维的向量空间$\mathrm{R}^K(K<L)$中。数据点$\boldsymbol{x}_i$在$\mathrm{R}^K$空间中表示为$\boldsymbol{z}_i$，且$\boldsymbol{z}_i=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i$。

判别邻域嵌入空间矩阵$\boldsymbol{A}$的训练过程包括以下步骤：

(1)给定具有说话人身份标注信息的训练数据集$\boldsymbol{X}$，第i个顶点代表第i条训练语句对应的i-vector向量$\boldsymbol{x}_i$，构建数据集上的判别邻接图G。

(2)计算第i个顶点和第j个顶点之间的亲合权重B_{ij}和重构权重W_{ij}，从而得到亲合权重矩阵$\boldsymbol{B}$和重构权重矩阵$\boldsymbol{W}$。

(a) B_{ij}反映了不同类数据对之间的亲合力。如果$\boldsymbol{x}_i$和$\boldsymbol{x}_j$属于不同说话人的语句，且$\boldsymbol{x}_i$是$\boldsymbol{x}_j$的k个最近邻之一(或者$\boldsymbol{x}_j$是$\boldsymbol{x}_i$的k个最近邻之一)，则$B_{ij}=\exp(||\boldsymbol{x}_i-\boldsymbol{x}_j||^2/t)$，其中参数$t$为适当常量。否则$B_{ij}=0$。

(b) W_{ij}反映了同类数据$\boldsymbol{x}_j$对重构$\boldsymbol{x}_i$时的贡献。如果$\boldsymbol{x}_i$和$\boldsymbol{x}_j$属于不同说话人的语句，则$W_{ij}=0$。如果$\boldsymbol{x}_i$和$\boldsymbol{x}_j$属于同一说话人的语句，可以通过最小化重构损失函数

$$\varphi(\boldsymbol{W})=\sum_{i=1}^{N}||\boldsymbol{x}_i-\sum\nolimits_j W_{ij}\boldsymbol{x}_j||^2 \tag{2}$$

其约束条件为$\sum_j W_{ij}=1$, $j=1, 2, \cdots, N$。求解出重构权重矩阵系数W_{ij}。

(3)计算DNE映射变换矩阵$\boldsymbol{A}$。DNE的思想是，在映射到低维空间后，类间样本彼此远离，类内样本局部结构保持不变，综合两方面考虑，给出以下2个最优化问题[20]

$$\max \sum_{i,j=1}^{N} ||\boldsymbol{y}_i - \boldsymbol{y}_j||^2 B_{ij} = \max(\boldsymbol{A}^{\mathrm{T}} \boldsymbol{X} \boldsymbol{H} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{A}) \tag{3}$$

$$\min \sum_{i=1}^{N} ||\boldsymbol{y}_i - \sum_{j} W_{ij} \boldsymbol{y}_j||^2 = \min(\boldsymbol{A}^{\mathrm{T}} \boldsymbol{X} \boldsymbol{M} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{A}) \tag{4}$$

约束条件为$\boldsymbol{A}^{\mathrm{T}} \boldsymbol{X} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{A}=1$。其中，式(3)中的$\boldsymbol{H}=\boldsymbol{D}-\boldsymbol{B}$，$\boldsymbol{D}$为对角阵，$D_{ii}=\sum_j B_{ij}$；式(4)中$\boldsymbol{M}=(\boldsymbol{I}-\boldsymbol{W})^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{W})$。

定义类间邻域散度$\boldsymbol{S}_B=\boldsymbol{X} \boldsymbol{H} \boldsymbol{X}^{\mathrm{T}}$，类内邻域散度$\boldsymbol{S}_W=\boldsymbol{X} \boldsymbol{M} \boldsymbol{X}^{\mathrm{T}}$，上述2个最优化问题可以转换成下面的广义特征值求解问题，即

$$(\boldsymbol{S}_B - \boldsymbol{S}_W)\boldsymbol{A}=\lambda \boldsymbol{A} \tag{5}$$

通过求解式(5)，可得到判别邻域嵌入空间矩阵$\boldsymbol{A}=[\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_K]$，其中，$\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_K$是上述问题的前$K$个最大特征值对应的特征向量。

4 实验

4.1 实验配置

本文使用NIST SRE 2010核心测试集(core-core)的电话训练、电话测试(tel-tel)作为测试集，采用等错误率(Equal Error Rate, EER)和最小检测错误代价(minimum Detect Cost Function, minDCF)对系统性能进行评价[21,22]。

实验中所使用的特征为36维的梅尔频率倒谱系数(Mel Frequency Cepstral Coefficents, MFCC)特征，其每帧特征由18维的基本倒谱系数及其1次差分构成，并使用特征弯折(feature warping)技术对特征进行规整。使用NIST SRE 2004 1side的目标说话人训练数据训练2个与性别相关的UBM，高斯数为1024。使用NIST SRE 2004, 2005以及2006的电话语音数据来训练总变化矩阵$\boldsymbol{T}$, NPE矩阵和DNE矩阵、WCCN转换矩阵以及LDA降维矩阵。选用NIST SRE 2004以及2005的部分数据来做负样本训练数据，使用SVMLight[23]来训练支持向量。

4.2 实验结果

本节在测试集上对本文提出的DNE算法和已有的NPE算法在说话人识别系统中进行了对比试验。

表1列出了在没有任何信道补偿技术下DNE算法和NPE算法的实验结果。从表1中可以看出，DNE的性能要优于NPE。相对于NPE, DNE在男声测试集上EER和minDCF分别有8.33%和5.39%的性能提升；在女声测试集上，EER和minDCF分别有9.03%和8.20%的性能提升。

表2是NPE算法以及本文提出的DNE算法在后端经过LDA之后的性能比较。从表2可以看到，当经过LDA信道补偿后，DNE系统性能优于NPE算法，在男声测试集上EER和minDCF分别有11.04%和7.93%的性能提升；在女声测试集上，EER和minDCF分别有8.84%和4.58%的性能提升。

表3是NPE算法以及DNE算法在后端经过WCCN信道补偿之后的性能比较。从实验结果我们可以看到，当经过WCCN信道补偿后，相对于NPE, DNE系统性能在男声测试集上EER和minDCF分别有9.47%和6.64%的性能提升；在女声测试集上，EER和minDCF分别有10.17%和8.86%的性能提升。

表4比较了NPE算法以及DNE算法在后端同时经过LDA和WCCN信道补偿之后的性能。从实验结果可以看到，当经过LDA和WCCN信道补偿后，DNE系统性能仍然优于NPE。相对于NPE, DNE系统性能在男声测试集上EER和minDCF分别有5.90%和8.82%的性能提升；在女声测试集上，EER和minDCF分别有8.39%和5.31%的性能提升。

表5将本文提出的DNE算法(经过LDA和WCCN信道补偿)与目前说话人识别领域的主流算法——概率线性判别分析(Probability Linear Discriminant Analysis, PLDA)[24,25]进行了性能对比，

表 1 NIST SRE 2010电话-电话测试集上DNE和NPE的EER和minDCF比较(无信道补偿)

系统	男声		女声	
	EER(%)	minDCF	EER(%)	minDCF
NPE	5.76	0.0575	6.98	0.0744
DNE	5.28	0.0544	6.35	0.0683

表 2 NIST SRE 2010电话-电话测试集上DNE和NPE的EER和minDCF比较(LDA信道补偿)

系统	男声		女声	
	EER(%)	minDCF	EER(%)	minDCF
NPE+LDA	4.71	0.0492	6.11	0.0633
DNE+LDA	4.19	0.0453	5.57	0.0604

表 3 NIST SRE 2010电话-电话测试集上DNE和NPE的EER和minDCF比较(WCCN信道补偿)

系统	男声		女声	
	EER(%)	minDCF	EER(%)	minDCF
NPE+WCCN	5.07	0.0512	6.49	0.0677
DNE+WCCN	4.59	0.0478	5.83	0.0617

表 4　NIST SRE 2010电话-电话测试集上DNE和NPE的EER和minDCF比较(LDA+WCCN信道补偿)

系统	男声		女声	
	EER(%)	minDCF	EER(%)	minDCF
NPE+LDA+WCCN	4.41	0.0476	5.72	0.0584
DNE+LDA+WCCN	4.15	0.0434	5.24	0.0553

表 5　NIST SRE 2010电话-电话测试集上DNE和PLDA的EER和minDCF比较

系统	男声		女声	
	EER(%)	minDCF	EER(%)	minDCF
DNE+LDA+WCCN	4.15	0.0434	5.24	0.0553
PLDA	4.12	0.0428	5.37	0.0532

本实验中采用的是高斯PLDA算法。从实验结果可以看到，DNE算法与PLDA算法性能相当。

5　结论

针对邻域保持嵌入技术的不足，本文将判别邻域嵌入算法引入到基于i-vector的说话人识别中。与邻域保持嵌入技术相比，判别邻域嵌入算法结合邻域和类的信息，对邻域作类别判断和区别对待，因此，能够有效地克服邻域保持嵌入技术存在的缺点，可以进一步提高说话人识别性能。

参 考 文 献

[1] REYNOLDS D A and ROSE R C. Robust text-independent speaker identification using Gaussian mixture speaker models[J]. *IEEE Transactions on Speech and Audio Processing*, 1995, 3(1): 72–83. doi: 10.1109/89.365379.

[2] KINNUNEN T and LI Haizhou. An overview of text-independent speaker recognition: From features to supervectors[J]. *Speech Communication*, 2010, 52(1): 12–40. doi: 10.1016/j.specom.2009.08.009.

[3] 王伟, 韩纪庆, 郑铁然, 等. 基于Fisher判别字典学习的说话人识别[J]. 电子与信息学报, 2016, 38(2): 367–372. doi: 10.11999/JEIT150566.

WANG Wei, HAN Jiqing, ZHENG Tieran, *et al*. Speaker recognition based on fisher discrimination dictionary Learning[J]. *Journal of Electronics & Information Technology*, 2016, 38(2): 367–372. doi: 10.11999/JEIT150566.

[4] KENNY P, BOULIANNE G, OUELLET P, *et al*. Speaker and session variability in GMM-based speaker verification[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2007, 15(4): 1448–1460. doi: 10.1109/tasl.2007.894527.

[5] 郭武, 戴礼荣, 王仁华. 采用因子分析和支持向量机的说话人确认系统[J]. 电子与信息学报, 2009, 31(2): 302–305. doi: 10.3724/SP.J.1146.2007.01289.

GUO Wu, DAI Lirong, and WANG Renhua. Speaker verification based on factor analysis and SVM[J]. *Journal of Electronics & Information Technology*, 2009, 31(2): 302–305. doi: 10.3724/SP.J.1146.2007.01289.

[6] 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. doi: 10.1109/tasl.2010.2064307.

[7] DHANUSH B K, SUPARNA S, AARTHY R, *et al*. Factor analysis methods for joint speaker verification and spoof detection[C]. Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing, New Orleans, USA, 2017: 5385–5389. doi: 10.1109/ICASSP.2017.7953185.

[8] SU Hang and WEGMANN S. Factor analysis based speaker verification using ASR[C]. Proceedings of the Interspeech 2016, San Francisco, USA, 2016: 2223–2227. doi: 10.21437/Interspeech.2016-1157.

[9] MAK M W, PANG Xiaomin, and CHIEN J T. Mixture of PLDA for noise robust i-vector speaker verification[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2016, 24(1): 130–142. doi: 10.1109/TASLP.2015.2499038.

[10] LEI Yun and HANSEN J H L. Speaker recognition using supervised probabilistic principal component analysis[C]. Proceedings of the Interspeech 2010, Makuhari, Japan, 2010: 382–385.

[11] LIANG Chunyan, YANG Lin, ZHAO Qingwei, *et al*. Factor Analysis of neighborhood-preserving embedding for speaker verification[J]. *IEICE Transactions on Information and Systems*, 2012, 95(10): 2572–2576. doi: 10.1587/transinf.e95.d.2572.

[12] YANG Jinchao, LIANG Chunyan, YANG Lin, *et al*. Factor analysis of Laplacian approach for speaker recognition[C]. Proceedings of 2012 IEEE International Conference on Acoustics, Speech and Signal Processing, Kyoto, Japan, 2012: 4221–4224. doi: 10.1109/ICASSP.2012.6288850.

[13] CHIEN J T and HSU C W. Variational manifold learning for speaker recognition[C]. Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing, New Orleans, USA, 2017: 4935–4939. doi: 10.1109/ICASSP.2017.7953095.

[14] WU Di. Speaker recognition based on i-vector and improved local preserving projection[C]. Proceedings of the 2015 Chinese Intelligent Automation Conference, Fuzhou, China, 2015: 115–121. doi: 10.1007/978-3-662-46469-4_12.

[15] HE Xiaofei, CAI Deng, YAN Shuicheng, *et al*. Neighborhood preserving embedding[C]. Proceedings of the

Tenth IEEE International Conference on Computer Vision, Beijing, China, 2005: 1208–1213. doi: 10.1109/ICCV.2005.167.

[16] KAJAREKAR S S and STOLCKE A. NAP and WCCN: Comparison of approaches using MLLR-SVM speaker verification system[C]. Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing, Honolulu, USA, 2007: IV-249–IV-252. doi: 10.1109/ICASSP.2007.366896.

[17] HAEB-UMBACH R and NEY H. Linear discriminant analysis for improved large vocabulary continuous speech recognition[C]. Proceedings of 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing, San Francisco, USA, 1992: 13–16. doi: 10.1109/ICASSP.1992.225984.

[18] DING Chuntao and ZHANG Li. Double adjacency graphs-based discriminant neighborhood embedding[J]. *Pattern Recognition*, 2015, 48(5): 1734–1742. doi: 10.1016/j.patcog.2014.08.025.

[19] WANG Jing, CHEN Fang, and GAO Quanxue. Discriminant neighborhood structure embedding using trace ratio criterion for image recognition[J]. *Journal of Computer and Communications*, 2015, 3(11): 61282. doi: 10.4236/jcc.2015.311011.

[20] 魏权龄, 王日爽, 徐冰, 等. 数学规划与优化设计[M]. 北京: 国防工业出版社, 1984: 358–470.
WEI Quanling, WANG Rishuang, XU Bing, *et al*. Mathematical Programming and Optimization Design[M]. Beijing: National Defense Industry Press, 1984: 358–470.

[21] NIST. The NIST year 2010 speaker recognition evaluation plan[EB/OL]. http://www.oalib.com/references/16891962, 2012.

[22] SCHEFFER N, FERRER L, GRACIARENA M, *et al*. The SRI NIST 2010 speaker recognition evaluation system[C]. Proceedings of 2011 IEEE International Conference on Acoustics, Speech and Signal Processing, Prague, Czech Republic, 2011: 5292–5295. doi: 10.1109/ICASSP.2011.5947552.

[23] JOACHIMS T. SVM-light support vector machine[EB/OL]. http://svmlight.joachims.org/, 2008.

[24] KINNUNEN T, JUVELA L, ALKU P, *et al*. Non-parallel voice conversion using i-vector PLDA: towards unifying speaker verification and transformation[C]. Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing, New Orleans, USA, 2017: 5535–5539. doi: 10.1109/ICASSP.2017.7953215.

[25] BAHMANINEZHAD F and HANSEN J H L. i-Vector/PLDA speaker recognition using support vectors with discriminant analysis[C]. Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing, New Orleans, USA, 2017: 5410–5414. doi: 10.1109/ICASSP.2017.7953190.

梁春燕：女，1986年生，讲师，研究方向为说话人识别、语种识别.

袁文浩：男，1985年生，讲师，研究方向为语音信号处理、语音增强.

李艳玲：女，1978年生，副教授，研究方向为自然语言处理、口语理解、机器学习.

夏 斌：男，1973年生，副教授，研究方向为深度学习、信号与信息处理.

孙文珠：男，1983年生，讲师，研究方向为多媒体信号传输.