Advances in Applied Mathematics 应用数学进展, 2020, 9(1), 30-37
Published Online January 2020 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2020.91004

Hans汉斯

Research on Small Sample Voiceprint Recognition Based on Deep Learning

Lv Han[1], Linhua Zhou[2], Wenlian Ma[1*], Weijie Zheng[2], Tao Ma[2], Tianxing Li[2]

[1]Department of Applied Mathematics, School of Science, Changchun University of Science and Technology, Changchun Jilin
[2]Provincial Demonstration Center for Experimental Mathematics Education (Changchun University of Science and Technology), Changchun Jilin
Email: *mawl@cust.edu.cn

Received: Dec. 12th, 2019; accepted: Dec. 28th, 2019; published: Jan. 3rd, 2020

Abstract

This paper studies the problem of small sample voiceprint recognition. In the experiment, the 39-dimensional features composed of the Mel cepstral coefficient and its dynamic differential coefficients are used as the basic acoustic features, and the basic acoustic features are extracted from the 128-dimensional depth acoustics through a deep belief network stacked by a three-layer restricted Boltzmann machine. Finally through the support vector machine and random forest for voiceprint recognition. Training deep belief networks, each speaker to choose short speech signal of small sample data as the network training set, trained deep belief network model at the same time as the depth of the acoustic feature extraction, with the characteristics of the extractor on the depth of the training focus on the speaker voice signal extraction acoustic characteristics, the generalization ability of the depth acoustic feature extractor is further verified. The experimental results show that the soundprint recognition model designed in this paper has high recognition accuracy, and the depth feature extractor has better generalization ability.

Keywords

Voiceprint Recognition, Deep Belief Networks, Support Vector Machine, Random Forest

基于深度学习的小样本声纹识别研究

韩　侣[1]，周林华[2]，马文联[1*]，郑伟杰[2]，马　涛[2]，李天星[2]

[1]长春理工大学理学院应用数学系，吉林 长春
[2]数学实验省级教学示范中心(长春理工大学)，吉林 长春
Email: *mawl@cust.edu.cn

收稿日期：2019年12月12日；录用日期：2019年12月28日；发布日期：2020年1月3日

*通讯作者。

文章引用: 韩侣, 周林华, 马文联, 郑伟杰, 马涛, 李天星. 基于深度学习的小样本声纹识别研究[J]. 应用数学进展, 2020, 9(1): 30-37. DOI: 10.12677/aam.2020.91004

摘　要

本文研究了小样本声纹识别问题。实验中采用梅尔倒谱系数与其动态差分系数组成的39维特征作为基本声学特征，再将基本声学特征通过由三层受限玻尔兹曼机堆叠而成的深度置信网络提取128维深度声学特征，最后通过支持向量机和随机森林进行分声纹识别。训练深度置信网络时，每个说话人选用短时语音信号组成的小样本数据作为该网络的训练集，同时将训练好的深度置信网络模型作为深度声学特征提取器，用该特征提取器对非训练集中说话人语音信号提取深度声学特征，进一步验证了该深度声学特征提取器的泛化能力。实验结果表明，本文设计的声纹识别模型识别准确率高，且深度特征提取器有较好的泛化能力。

关键词

声纹识别，深度置信网络，支持向量机，随机森林

1. 引言

声纹识别(Voiceprint Recognition, VPR)是通过分析每个说话人声纹特征之间的差异来达到对未知语音进行识别的目的，因此声纹识别又称为说话人识别。声纹识别研究的重点是声学特征提取和声学特征建模，在声学特征提取方面常用特征有线性预测倒谱系数[1] (Linear Predictive Cepstrum Coefficient, LPCC)、感知线性预测系数[2] (Perceptual Linear Predictive, PLP)以及梅尔倒谱系数[3] [4] (Mel-frequency Cepstrum Coefficient, MFCC)。在声学特征建模上相继采用了矢量量化法[5] [6] (Vector Quantization, VQ)、隐马尔科夫模型[7] (Hidden Markov Model, HMM)、人工神经网络[8] (Artificial Neural Network, ANN)以及高斯混合模型－通用背景模型[9] (Gaussian mixture model-universal background model, GMM-UBM)等技术。

在深度学习用于声纹识别之前，GMM-UBM 是声纹识别广泛应用的技术之一，深度学习技术既可以作为一个深度声学特征提取器，同时也可以作为分类器[10] [11]，如田垚等人采用深度信念网络在基本声学特征(MFCC)的基础上进行瓶颈特征提取并结合高低混合模型等模型进行识别[12]，并通过实验证明了较传统的声学特征而言，该方法具有一定的优势，闫河等人将信号频谱图通过卷积神经网络进行声纹识别[13]，约翰斯霍普金斯大学的 Povey 提出基于 DNN 的 x-vector 说话人确认系统，该系统将语音特征提取过程分为帧级(frame-level)和段级(segment-level)，并使用统计池化层连接两级特征[14]。虽然深度学习在特征提取是信号匹配上都有较好的表现，但模型的训练过程中需要大量的训练数据，而且在实际应用中不可能每增加新的说话人就从新训练一个深度网络模型，非训练集中的说话人通过这个深度网络所提取特征的表征能力也需进一步实验，即模型的泛化能力也需进一步验证。

对于上述问题，本文针对短时语音信号，采用小样本进行深度模型训练，并针对非训练集中的说话人的语音信号，也通过这个深度网络模型进行深度声学特征提取，最后通过支持向量机和随机森林进行分类验证，实验结果表明，在本文构建的深度网络结构提取的深度声学特征能在有监督分类器上得到较好的识别率。

2. 深度学习方法概述

在声纹识别中常用的深度学习方法有多层感知器、卷积神经网络以及深度置信网络，其深度体现在隐含层的层数上，深度神经网络使用更多的隐含层的目的是希望更深层的学习后能对数据有更强的表征能力。深度信念网络 DBNs 是深度神经网络 DNN 的一种，在语音识别领域取得很大的成功，该网络体现了无监督学习在各层训练的有效性，指出各层可在前一层训练结果输出的基础上，再次进行无监督训练，在预训练的基础上模型获得了较好的初始参数，最后在通过反向传播算法对模型的参数进行反向调优。

2.1. 基于深度置信网络的参数预训练

受限玻尔兹曼机(restricted Boltzmann machines, RBM)是一类具有两层结构、对称连接且无反馈的神经网络模型，层间全连接，层内无连接。假设一个 RBM 有 n 个可见单元和 m 个隐单元，用向量 v 和 h 表示可见单元和隐单元的状态，其中 v_i 表示第 i 个可见单元的状态，h_j 表示第 j 个可见单元状态，那么，对于一组给定的状态 (v,h)，RBM 作为一个系统所必备的能量定义为[15]：

$$E(v,h|\theta)=-\sum_{i=1}^{n}\alpha_i v_i-\sum_{j=1}^{m}b_j h_j-\sum_{i=1}^{n}\sum_{j=1}^{m}v_i W_{ij}h_j \text{ 。}$$

式中：W_{ij} 为权重，表示连接了可见单元 i 与隐单元 j，α_i 与 b_j 分别表示可见单元 i 与隐单元 j 的偏置，当参数确定时，基于能量函数，可以得到 (v,h) 的联合概率分布

$$P(v,h|\theta)=\frac{\mathrm{e}^{-E(v,h|\theta)}}{Z(\theta)},\ Z(\theta)=\sum_{v,h}\mathrm{e}^{-E(v,h|\theta)} \text{ 。}$$

式中 $Z(\theta)$ 为归一化因子，由 RBM 所定义的关于观测数据 v 的分布，即联合概率分布的边界分布，也称似然函数(likelihood function)，RBM 的结构为层间连接，层内无连接，当给定可见单元的状态时，各隐单元与可见单元之间的激活状态是条件独立的，此时，可见单元与隐单元的激活概率为：

$$p(h_j=1|v,\theta)=\sigma\left(b_j+\sum_j v_i W_{ij}\right) \text{ 。}$$

$$p(v_i=1|h,\theta)=\sigma\left(a_i+\sum_i W_{ij}h_j\right) \text{ 。}$$

式中 $\sigma(x)=\dfrac{1}{1+\exp(-x)}$ 为 sigmoid 激活函数。

RBM 的优化目标是要最大化可见层节点概率分布，在训练过程中可以通过对比散度算法(contrastive divergence, CD)来得到模型中的参数。对比散度算法的输入是一个训练样本 x_0，设隐层单元个数 m，学习率 ϵ，最大训练周期 T，输出是连接权重矩阵 W，可见层偏置 a，隐藏层偏置向量 b，算法描述如下：

令可见层单元的初始状态 $v_1=x_0$，W,a,b 为 0~1 之间随机数。

For $t=1,2,\cdots,T$；

For $j=1,2,\cdots,m$ (对所有隐单元)；

计算 $P(h_{1j}=1|v_1)$，即 $P(h_{1j}=1|v_1)=\sigma\left(b_i+\sum_i v_{1i}W_{ij}\right)$；

从条件分布 $P(h_{1j}|v_1)$ 中抽取 $h_{1j}\in\{0,1\}$ 。

End For

For $i=1,2,\cdots,n$ (对所有可见单元)；

计算 $P(v_{2i}=1|h_1)$，即 $P(v_{2i}=1|h_1)=\sigma\left(a_i+\sum_j W_{ij}h_{1j}\right)$；

从条件分布 $P(v_{2i}\mid h_1)$ 中抽取 $v_{2i}\in\{0,1\}$。

End For

For $j=1,2,\cdots,m$ (对所有隐单元)；

计算 $P(h_{2j}=1\mid v_2)$，即 $P(h_{2j}=1\mid v_2)=\sigma\left(b_i+\sum_i v_{2i}W_{ij}\right)$。

End For

最后参数的更新计算为：

$$W\leftarrow W+\epsilon\left(P(h_1=1\mid v_1)v_1^{\mathrm{T}}-P(h_2=1\mid v_2)v_2^{\mathrm{T}}\right)$$
$$a\leftarrow a+\epsilon(v_1-v_2)\quad 。$$
$$b\leftarrow b+\epsilon\left(P(h_1=1\mid v_1)-P(h_2=1\mid v_2)\right)$$

2.2. 基于反向传播算法的参数调优

当 DBN 完成预训练之后，需要经过基于反向传播算法的参数调优。首先将 DBN 各层的网络参数做为 DNN 的参数的初值，最后加上一层 softmax 函数层得到完整的 DNN 结构。Softmax 层的每个输出节点对应一个类别的概率，从而实现分类的目的。

假设最终的 DBN 结构有 N 个 RBM 构成，则这个结构共有 $N+1$ 层，该网络结构中第 1 层表示输入层，第 $N+1$ 层表示输出层，隐藏层 k 的取值为 $k=2,3,\cdots,N$。设 ω^k 和 b^k 表示隐藏层的权重和偏置，v^{k+1} 为上一层输入的加权，激活函数 σ 通常可以选取 sigmoid、tanh 和 ReLu 等函数。其隐藏层节点的输出可以表示为：

$$v^{k+1}=\omega^k h^k+b^k\quad h^k=\sigma\left(v^k\right)。$$

对于输出层通常采用 softmax 函数：

$$p_s=\frac{\exp\left(v_s^{N+1}\right)}{\sum_j\exp\left(v_s^{N+1}\right)}。$$

其中：j 为输出类别索引，p_s 表示第 p_s 类输出类别的概率分值。在使用 BP 算法反向传播的过程中，通常使用交叉熵作为损失函数，通过最小化代价函数来修正 DBN 结构中的参数，d_s 是一个维的向量，s 代表正确的类别值，当 s 为训练数据所属类别时 $d_s=1$，当 s 为其它类别时 $d_s=0$，交叉熵函数为：

$$L=-\sum_s d_s\log p_s。$$

3. 有监督分类器

3.1. 支持向量机

支持向量机(Support Vector Machine, SVM)是在统计学习理论和结构风险最小原理基础上发展起来的一种学习方法，其机理可以简单地描述为：寻找一个满足分类要求的分割超平面，使训练集中的点距离该分割超平面尽可能地远，即寻找一个最优分割超平面，使其两侧的空白区域最大。SVM 在解决小样本、非线性和高维模式识别问题中表现出特有的优势，并在很大程度上克服了“维数灾难”和“过学习”等问题，并有泛化能力强等优点。

对于给定样本集 $\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$，其中 $x_i\in R^d$，$y_i\in\{\pm1\}$，支持向量机的目标是构造出一个超平面 $\omega\cdot x+b=0$ 将两类不同样本分割，使得两类间隔最大。相应的分类决策函数为[16]：

$$f(x)=\omega x+b。$$

若 $f(x)>0$，则待测样本分为一类；若 $f(x)<0$，待测样本分为另一类。

假设给定的样本是线性可分的，则求最优分类面即求解二次规划问题：

$$\begin{aligned}&\text{Minimize } \Phi(\omega,b)=\frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{n}\xi_i \\ &\text{s.t. } y_i(\omega\cdot x+b)-1+\xi_i\geq 0,\ i=1,2,\cdots,n\end{aligned}。$$

使用 Lagrange 优化算法可将上述最优分类面问题转化为对偶问题，并得到最优分类决策函数：

$$f(x)=\operatorname{sgn}\{(w\cdot x)+b\}=\operatorname{sgn}\left\{\sum_{i=1}^{m}\alpha_i^* y_i(x_i\cdot x)+b^*\right\}。$$

其中，$\alpha^*=(\alpha_1^*,\alpha_2^*,\cdots,\alpha_n^*)$ 为拉格朗日乘子，$b^*=\frac{1}{2}\left[(w^*\cdot x(1))+(w^*\cdot x(-1))\right]$ 是分类阈值，$x(1)$ 表示两类中属于第一类的任意支持向量，$x(-1)$ 表示属于第二类的任意支持向量，w^* 为权值。

对于非线性问题，通常是通过核函数 $K(x,x')$ 将原始数据映射到高维的特征空间中，转化为高维空间中的线性问题，在变换后的空间中求最优分类超平面，即分类函数变为：

$$f(x)=\operatorname{sign}\left(\sum_{i=1}^{n}\alpha_i^* y_i K(x_i\cdot x)+y_j-\sum_{i=1}^{n}\alpha_i^* y_i K(x_i\cdot x_j)\right)。$$

常用的核函数主要有线性核函数、多项式核函数以及高斯核函等。

3.2. 随机森林

随机森林是一个包含多个决策树的分类器，是一种重要的基于 Bagging 的集成学习方法，可以用作分类和回归。由于采用有放回的采样来构造不同数据集训练模型，模型的泛化能力通常比单一模型强。随机森林训练多个决策树弱分类器，然后组合这些弱分类器形成一个强分类器，通过投票的方式来得到最终的分类结果。和普通的决策树不同，随机森林在训练决策树时并不是选择一个全局最优的特征来分裂节点，而是先随机选择部分样本特征，然后再到里面选择一个局部最优的特征来为决策树划分左右子树，进一步提升了模型的泛化能力。

传统决策树在选择最优属性时是对所有可用特征进行选择，而在随机森林(Random Forest, RF)，对每一个基决策树，先从所有特征中随机抽样出一个包含 k 个特征的子集，然后基于该子集进行属性划分[17]。具体算法流程如下：

输入：训练数据集 D，弱学习器算法 G，弱学习器迭代次数 T；

输出：强学习器 $H(x)$。

Step 1：对于 $t=1,2,\cdots,T$：

1) 对训练数据集 D 采用自助法采样，得到包含 m 个样本的采样集 D_t。

2) 使用采样集 D_t 训练第 t 个决策树模型 $G_t(x)$，从所有特征中随机抽样出一个包含 k 个特征的子集，使用 CART 算法构建决策树。

Step 2：若为分类问题，则使用 T 个模型的类别标签进行投票预测；若为回归问题，则用简单平均法进行预测。

随机森林算法结构简单、易于实现且计算开销较小，在很多现实任务中展现出强大的性能。相比于 Bagging，随机森林不仅继承了通过样本扰动带来的样本多样性，还引入属性扰动进一步提升了模型的泛化能力。尽管随机森林中个体学习器的性能往往有所下降，但随着个体学习器数量的增多，模型的整体性能会获得更大的提升。同时，随机森林的训练效率往往优于 Bagging，因为随机森林的个体学习器在进行属性划分时需要计算的特征个数更少。

4. 实验设置与结果分析

4.1. 声学特征提取

4.1.1. 实验数据设置及 MFCC 提取

本文采用 AISHELL-ASR0009-OS1 语音数据库，音频降采样为 16 kHz 数据库中包含 400 名来自中国不同口音区域的发言人。在实验设置中，根据不同的实验条件将划分不同的训练集与测试集。

梅尔到谱系数的提取包含预滤波、预加重、分帧、加窗、快速傅立叶变换、三角窗滤波、求对数、离散余弦变换、倒谱均值减、差分等步骤，在本实验中，设置帧长为 25 ms、帧移为 10 ms，滤波器组的滤波器数量为 26，预加重过滤器的系数为 0.97，最后得到的倒频谱数量为 13，即 13 维 MFCC，最后将 13 维 MFCC 特征与其一阶差分与二阶差分组合，得到 39 维的基本声学特征。若假设某人某句话共有 n 帧，每帧提取的 MFCC 记为 $X_k\left(k=1,2,\cdots,n\right)$，则这句话的 MFCC 重新计算为：$\frac{1}{n}\sum_{k=1}^{n}x_k$。

4.1.2. 深度声学特征 d-vector

深度置信网络由 RBM 堆叠组成，每个 RBM 的权值利用吉布斯采样进行估计，本实验的 DBN 采用三层 RBM (网络节点为：39-128-128-128)堆叠，如下图 1 所示，最终以最后一层 128 个节点的输出值作为由 39 维 MFCC 经过 DBN 进行特征再提取得到深度声学特征。

4.2. 基于深度学习的声纹识别结果

本文实验数据来自于 AISHELL-ASR0009-OS1 语音数据库，从中选择一定的数据并设置了两个数据集，分别为每人 5 条语音信号和每人 10 条语音数据作为深度神经网络的训练集，每个数据集中共 200 人，具体设置如下表 1 所示。

Table 1. Experimental data set settings
表 1. 实验数据集设置

数据集	实验人数	训练集语料数	测试集语料数
数据集 1	200	5	50
数据集 2	200	10	50

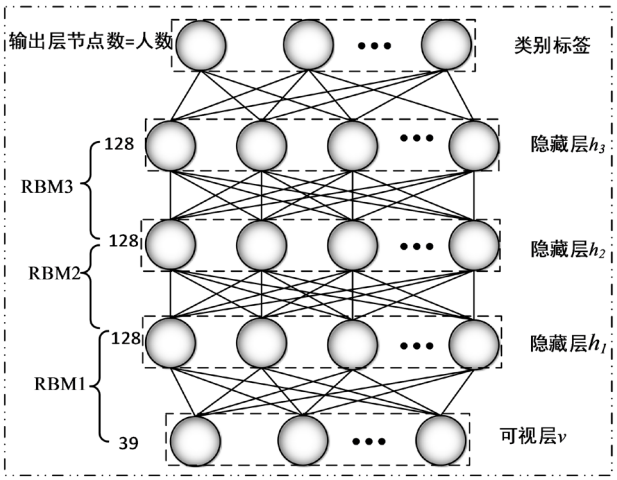

Figure 1. DBN structure
图 1. DBN 结构

在本实验中首先采用支持向量机和随机森林模型对基本声学特征 MFCC 进行识别，然后在通过以上模型对深度声学特征 d-vector (128 dim)进行对比。

对以上两种声纹特征识别的模型中，模型采用相同的参数，如在随机森林和决策树中，树的深度均采用 15，支持向量机中的核函数均采用线性核函数，得到的模型的准确率，召回率如下表 2 与表 3 所示。

Table 2. Recognition rate of MFCC (39 dim) in the model
表 2. MFCC (39 dim)在模型中的识别率

模型	数据集 1 准确率	数据集 2 准确率
随机森林	46.78%	60.08%
SVM	68%	71.22%

Table 3. Recognition rate of d-vector (128 dim) in the model
表 3. d-vector (128 dim)在模型中的识别率

模型	数据集 1 准确率	数据集 2 准确率
随机森林	75.21%	90.19%
SVM	86.8%	95.6%

4.3. 深度神经网络模型泛化验证

在本文所设置的实验中，主要分为基本声学特征 MFCC 提取、深度声学特征 d-vector 提取和有监督分类器的训练三个步骤，其中在深度特征提取这一阶段主要是通过训练一个深度神经网络，最终选用该网络中某一层的节点作为深度声学特征。

在模型的训练时分为深度神经网络训练和分类器的模型训练，深度特征的提取主要在于将低维基本声学特征变为高维的深度特征，此深度神经网络是采用了深度信念网络结构，该网络的特点是先进行无监督进行预训练，再进行有监督的微调，在这个过程中将语音信号中包括信道信息、语义信息、情绪以及说话人身份信息进行放大，在微调过程中是通过身份信息进行反向传播，这使得说话人身份信息相较于其它信息得以放大，因此深度特征在支持向量机等模型上进行分类时表现更好。

为了验证深度特征提取器的泛化能力，选取一定的语料设置如下的训练集与测试集，训练集中人数设置为 150 人，每人 10 条语料参与深度网络训练，其中网络结构以及分类器参数设置均与前实验相同，参与训练与未参与训练测试集每人均 50 条语料，实验结果如下表 4 所示。

Table 4. Model generalization capability verification results
表 4. 模型泛化能力验证结果

语料集	人数	训练集每人语料数	测试每人集语料数	SVM Accuracy	DF Accuracy
参与训练的话者	150	10	50	95.92%	92.01%
未参与训练话者	50		50	91.5%	90.2%

从表 4 可知，在参与深度信念网络训练的数据集上，支持向量机与随机森林的准确率分别为 95.92%和 92.01%，而在未参与训练的数据集上支持向量机与随机森林的准确率分别为 91.5%和 90.2%。从实验结果可知，虽然这 50 人没有参与神经网络模型的训练，但是通过该深度声学特征提取器提取的深度声学特征在有监督分类其上任有较好的表征能力，故深度声学特征提取器拥有较好的泛化能力，因此实际应用中，在一个集合中增加说话人的情况下，运用该网络结构不需从新训练深度网络模型。

5. 结论

本文运用了深度学习的方法构建了一个基于声纹识别的深度声学特征特征提取器，并将该特征结合支持向量机和随机森林等分类器进行声纹识别研究。实验中对比了基本声学特征 MFCC 与高维深度声学特征 d-vector 在支持向量机和树模型上的识别结果，实验结果表明支持向量机在两种特征上的分类性能更好，同时也对深度声学特征提取器的泛化能力进行了验证，并取得了较好的分类结果

参考文献

[1] Atal, B.S. (1976) Automatic Recognition of Speakers from Their Voices. *Proceedings of the IEEE*, **64**, 460-475. https://doi.org/10.1109/PROC.1976.10155

[2] Hermansky, H. (1990) Perceptual Linear Predictive (PLP) Analysis of Speech. *The Journal of the Acoustical Society of America*, **87**, 1738-1752. https://doi.org/10.1121/1.399423

[3] Davis, S.B. (1980) Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. *IEEE Transactions on Acoustics*, *Speech*, *and Signal Processing*, **28**, 65-74. https://doi.org/10.1109/TASSP.1980.1163420

[4] Furui, S. (1981) Cepstral Analysis Technique for Automatic Speaker Verification. *IEEE Transactions on Acoustics Speech and Signal Processing*, **29**, 254-272. https://doi.org/10.1109/TASSP.1981.1163530

[5] Burton, D.K. (1987) Text-Dependent Speaker Verification Using Vector Quantization Source Coding. *IEEE Transactions on Acoustics Speech and Signal Processing*, **35**, 133-143. https://doi.org/10.1109/TASSP.1987.1165110

[6] Soong, F.K. and Rosenberg, A.E. (1988) On the Use of Instantaneous and Transitional Spectral Information in Speaker Recognition. *IEEE Transactions on Acoustics Speech & Signal Processing*, **36**, 871-879. https://doi.org/10.1109/29.1598

[7] Naik, J.M., Netsch, L.P. and Doddington, G.R. (1989) Speaker Verification over Long Distance Telephone Lines. *International Conference on Acoustics*, *Speech*, *and Signal Processing*, Glasgow, 23-26 May 1989, 524-527.

[8] Yang, Y., Ren, W., Hui, Z., *et al*. (2012) The Research of Voiceprint Recognition Based on Genetic Optimized RBF Neural Networks. *IEEE International Conference on Computer Science & Automation Engineering*, Zhangjiajie, 25-27 May 2012, 425-428. https://doi.org/10.1109/CSAE.2012.6272630

[9] Abu, M.A., Zakariya, Q., *et al*. (2018) New Transformed Features Generated by Deep Bottleneck Extractor and a GMM-UBM Classifier for Speaker Age and Gender Classification. *Neural Computing & Applications*, **30**, 2581-2593. https://doi.org/10.1007/s00521-017-2848-4

[10] Variani, E., Lei, X., McDermott, E., Lopez Moreno, I. and Gonzalez-Dominguez, J. (2014) Deep Neural Networks for Small Footprint Text-Dependent Speaker Verification. *IEEE International Conference on Acoustics*, *Speech and Signal Processing*, Florence, 4-9 May 2014, 4052-4056. https://doi.org/10.1109/ICASSP.2014.6854363

[11] Hinton, G., Deng, L., Yu, D., *et al*. (2012) Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Processing Magazine*, **29**, 82-97. https://doi.org/10.1109/MSP.2012.2205597

[12] 田垚, 蔡猛, 何亮, 刘加. 基于深度神经网络和 Bottleneck 特征的说话人识别系统[J]. 清华大学学报(自然科学版), 2016, 56(11): 1143-1148.

[13] 闫河, 董莺艳, 王鹏, 罗成, 李焕. 基于 CNN-LSTM 网络的声纹识别研究[J]. 计算机应用与软件, 2019, 36(4): 166-170.

[14] Snyder, D., *et al*. (2017) Deep Neural Network-Based Speaker Embeddings for End-to-End Speaker Verification. *Spoken Language Technology Workshop*, San Diego, 13-16 December 2016, 165-170.

[15] 张春霞, 姬楠楠, 王冠伟. 受限玻尔兹曼机[J]. 工程数学学报, 2015, 32(2): 161-175.

[16] Hearst, M.A. (1998) Support Vector Machines. *IEEE Intelligent Systems & Their Applications*, **13**, 18-28. https://doi.org/10.1109/5254.708428

[17] Ho, T.K. (1995) Random Decision Forests. *International Conference on Document Analysis & Recognition*, Montreal, 14-16 August 1995, 278-282.