不相关匹配追踪的分段区分性特征变换方法

陈 斌[1,2],牛 铜[1],张连海[1],屈 丹[1],李弼程[1]
(1. 解放军信息工程大学信息系统工程学院,河南郑州 450001;2. 西南电子电信技术研究所上海分所,上海 200434)

摘 要: 为了提高基于分帧特征变换方法的稳定性,提出了一种基于分段的区分性特征变换方法.该方法将特征变换当成高维信号的稀疏逼近问题,采用状态绑定的方法训练得到基于域划分的线性变换矩阵(Region Dependent Linear Transform,RDLT)和基于最小音素错误准则均值补偿的特征(mean-offset feature Minimum Phone Error,m-fMPE)变换矩阵,将两者的特征变换矩阵构成过完备的字典;采用强制对齐的方式对语音信号进行分段,以似然度最大化作为目标函数,利用匹配追踪算法对目标函数迭代优化,自动地确定各语音信号段中的变换矩阵及其系数.为保证特征变换的稳定性,在选择变换矩阵过程中引入相关度测量,去除相关的特征基矢量.实验结果表明,相比于传统的 RDLT 方法,当声学模型分别采用最大似然和区分性准则训练时,识别性能分别可以提高 1.63% 和 2.23%.该方法同时能应用于语音增强和模型区分性训练中.

关键词: 特征变换;语音识别;区分性训练;语音增强;匹配追踪
中图分类号: TN912 **文献标识码:** A **文章编号:** 0372-2112 (2016)12-2924-08
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2016.12.016

A Discriminative Segmental Feature Transform Method Based on Uncorrelated Matching Pursuit

CHEN Bin[1,2], NIU Tong[1], ZHANG Lian-hai[1], QU Dan[1], LI Bi-cheng[1]
(1. *Institute of Information System Engineering, Information Engineering University, Zhengzhou, Henan* 450001, *China*;
2. *Shanghai Branch of Southwest Electronics and Telecommunication Technology Research Institute, Shanghai* 200434, *China*)

Abstract: A discriminative segmental feature transform method is proposed to promote the stability of the frame based method. The feature transform is considered as the sparse high dimensional approximation problem. Firstly, a set of feature transform matrices are estimated by tied-state based training of RDLT (Region Dependent Linear Transform) and m-fMPE (mean-offset feature Minimum Phone Error), and the transform matrices are integrated into an over-complete dictionary. Then, the speech signal is segmented through force alignment. Finally, following the matching pursuit to optimize the likelihood objective function iteratively, the transform matrices of each segment are selected from the dictionary and the corresponding coefficients are automatic determined in the optimization process. Further, to guarantee the stability of the transform matrices, a correlation measurement is introduced to remove the correlated basis in the recurrence process. The experimental results show that, compared with the traditional RDLT method, when the acoustic model is trained with maximum likelihood and discriminative training criterion separately, the recognition performance can be improved by 1.63% and 2.23% respectively. The method can also be applied to speech enhancement and model discriminative training.

Key words: feature transform; speech recognition; discriminative training; speech enhancement; matching pursuit

1 引言

目前,主流语音识别系统中常对识别特征进行特征变换[1,2],以进一步得到具有鲁棒性和区分性的特征.其中,采用高斯混合模型(Gaussian Mixture Model, GMM)进行声学空间划分的特征变换方法应用较为广泛,如基于最小音素错误准则的特征变换(feature Minimum Phone Error, fMPE)[3]和基于域划分的线性特征变换(Region Dependent Linear Transform, RDLT)[4~6].在此基础上,陆续提出了结合高斯混元参数信息的均值补

收稿日期:2015-05-17;修回日期:2015-11-24;责任编辑:覃怀银
基金项目:国家自然科学基金(No.61175017, No.61403415);国家高技术研究发展计划(863 计划)课题(No.2012AA011603)

偿(mean-offset) m-fMPE[7] 方法和状态绑定的(tied-state) RDLT[8] 方法,并同时应用于深度神经网络(Deep Neural Network,DNN)[9,10] 中,通过调整网络权值进行特征变换[11~13].

上述区分性特征变换方法中,训练阶段均是采用一段有限长信号求取变换矩阵,而在测试阶段却是对每一帧信号进行特征变换和补偿,这易造成训练和识别间不匹配. 另外,由于语音信号具有短时平稳性,一帧信号往往较难得到稳定的参数信息.

为了有效地解决不匹配问题,得到稳定的解. 在测试阶段,本文同样基于一段信号进行特征变换,即根据信号段的统计量信息,在训练得到的变换矩阵集合中,自动地选择特征变换矩阵. 在这个过程中变换矩阵个数的选取是关键,当选择的变换矩阵较少时,将不能得到精确的变换参数;而当选择的矩阵过多时,会使得特征参数的稳健性不够. 由于一次求解过程拥有的数据量有限,所选择的特征变换矩阵数相比于变换矩阵集合很小,是一个稀疏逼近问题.

本文将压缩感知理论引入到区分性特征变换中,在对语音信号分段的基础上,基于每一语音段求解其特征变换矩阵. 先采用状态绑定的方式训练得到变换矩阵,结合 RDLT 特征变换矩阵和均值补偿 fMPE 偏移矢量构成过完备字典,在特征域进行特征变换相关参数的稀疏表示,利用匹配追踪算法自动地确定变换矩阵个数及其系数,得到最终的变换矩阵. 为了保证变换矩阵的稳定性,在变换矩阵的选取过程中要求特征基矢量间不相关,并进一步讨论了不同分段方法对识别结果的影响.

2　基于语音分段的区分性特征变换

本文先采用状态绑定的方法得到 RDLT 变换矩阵和均值补偿 fMPE 偏移矢量,组成变换矩阵和偏移矢量集合,在此基础上结合压缩感知方法,采用最大似然准则进行特征变换矩阵和偏移矢量的选取.

2.1　基于状态绑定的特征变换矩阵

2.1.1　基于域划分的特征变换矩阵

RDLT[5] 利用全局的 GMM 模型将声学空间分成多个域,每个高斯混元对应一个域划分,通过区分性训练得到一个变换矩阵集合,每个变换矩阵对应于声学空间中的一个域. 用特征向量所属域对应的变换矩阵对其进行变换,特征所属的域由其在高斯混元的后验概率所决定,最终特征变换式(1)所示:

$$F_{\mathrm{RDLT}}(\boldsymbol{o}(t)) = \sum_{i=1}^{R} \kappa_t^{(i)} \boldsymbol{A}_i \boldsymbol{o}(t) \tag{1}$$

其中,$\boldsymbol{o}(t)$为时刻 t 的输入特征,声学空间共划分为 R 个域,$\boldsymbol{A}_i$ 为第 i 个域对应的变换矩阵,$\kappa_t^{(i)}$ 为 $\boldsymbol{o}(t)$属于第 i 个域的概率,可采用 GMM 混元后验概率来表示. 通常,RDLT 方法中变换矩阵 $\boldsymbol{A}_i$ 基于词图信息(lattice),根据 MPE 准则更新,声学模型参数通过最大似然(Maximum Likelihood,ML)准则更新. 这里采用状态绑定的方式求解 $\boldsymbol{A}_i$.

2.1.2　基于最小音素错误准则的特征变换

fMPE[3] 方法将特征在高斯混元上的后验概率组成一个新特征,将这个特征映射为一个偏移矢量,加在原始特征上. fMPE 方法中每个域对应一个偏移矢量,由于偏移矢量所含的信息量有限,常通过采用增大域的个数来保证其性能. 而 m-fMPE[7] 通过加入所在域的高斯混元参数信息,进而提高了每一个域中的信息量,m-fMPE 其变换式(2)所示:

$$F_{\text{m-fMPE}}(\boldsymbol{o}(t)) = \boldsymbol{o}(t) + \boldsymbol{M}\boldsymbol{h}_t \tag{2}$$

其中,$\boldsymbol{h}_t$ 由后验概率向量 $\boldsymbol{\kappa}_t$ 和均值补偿向量 $\boldsymbol{\delta}_t$ 组成,需要求取变换矩阵 $\boldsymbol{M}$.

$$\begin{aligned}
&\boldsymbol{h}_t = [\eta\boldsymbol{\kappa}_t, \boldsymbol{\delta}_t]^{\mathrm{T}} \\
&\boldsymbol{\delta}_t = [\kappa_t^{(1)}\boldsymbol{d}_{(t,1)}^{\mathrm{T}}, \kappa_t^{(2)}\boldsymbol{d}_{(t,2)}^{\mathrm{T}}, \cdots, \kappa_t^{(R)}\boldsymbol{d}_{(t,R)}^{\mathrm{T}}] \\
&\boldsymbol{d}_{(t,i)} = \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{o}(t) - \boldsymbol{\mu}_i)
\end{aligned} \tag{3}$$

式中,$\boldsymbol{\mu}_i$ 和 $\boldsymbol{\Sigma}_i$ 分别为第 i 个高斯分量的均值和协方差矩阵,$\kappa_t^{(i)}$ 为 $\boldsymbol{\kappa}_t$ 的第 i 个元素. fMPE 和 m-fMPE 的不同点在于向量 $\boldsymbol{h}_t$ 中是否含有 $\boldsymbol{\delta}_t$ 向量. 可进一步将变换矩阵 $\boldsymbol{M}$ 拆分为关于变量 $\boldsymbol{o}(t)$的变换矩阵 $\boldsymbol{H}$ 和偏移矢量 $\boldsymbol{b}$,即:

$$\begin{aligned}
&F_{\text{m-fMPE}}(\boldsymbol{o}(t)) \\
&= \boldsymbol{o}(t) + \sum_{i=1}^{L} (\kappa_t^{(i)}\boldsymbol{M}_a^{(i)} d_{(t,i)} + \kappa_t^{(i)} M_b^{(i)}) \\
&= \sum_{i=1}^{L} \kappa_t^{(i)} [(\boldsymbol{I} + \boldsymbol{M}_a^{(i)}\boldsymbol{\Sigma}_i^{-1})\boldsymbol{o}(t) + (\boldsymbol{M}_b^{(i)} - \boldsymbol{M}_a^{(i)}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i)] \\
&= \sum_{i=1}^{L} \kappa_t^{(i)} (\boldsymbol{H}_i\boldsymbol{o}(t) + \boldsymbol{b}_i)
\end{aligned} \tag{4}$$

其中,$\boldsymbol{M}_a$ 和 $\boldsymbol{M}_b$ 分别为 m-fMPE 均值补偿向量和后验概率向量所对应的变换矩阵,L 是声学空间的域划分个数. 基于状态绑定的 RDLT 和 m-fMPE 的求解过程相类似,只是求微分时针对的变量不同,以及确定迭代步长时有所差异,这里根据文献[8]分别进行求解.

2.2　基于分段区分性特征变换的一般形式

不同于传统方法中先验地设定所需变换矩阵的个数,再根据后验概率值的大小进行选择和加权. 这里先对语音信号进行分段,对每一语音段根据其声学统计量信息,利用最大似然准则,采用一种可变变换矩阵个数的方式,得到区分性特征变换的一般表达式.

2.2.1　基于变换矩阵字典的特征变换

设经过域划分后总共有 R 个域,每一个域对应的变换矩阵为 $\boldsymbol{A}_i$,语音信号被分成 S 段,其中第 s 个语音段的特征变换可以描述为式(5):

$$\boldsymbol{o}'^{s}(t)=\sum_{i=1}^{R}x_i^s\boldsymbol{A}_i\boldsymbol{o}^s(t) \tag{5}$$

式中,x_i^s 为所选择的特征变换矩阵 $\boldsymbol{A}_i$ 对应的权重系数,由于以下论述中,均在语音段 s 内求解相关参数,为了行文的简化,将上标 s 略去. 依据最大似然准则和期望最大(Expectation Maximization, EM)算法,需要最大化变换后特征的似然度[14,15],其目标函数为式(6):

$$\begin{aligned}\hat{\boldsymbol{x}}&=\arg\max_{x}\sum_{t=1}^{T}\sum_{m=1}^{M}\gamma_m(t)\log p(\boldsymbol{o}'(t)|\boldsymbol{\mu}_m)\\&=\arg\max_{x}\Big\{-\frac{1}{2}\sum_{t=1}^{T}\sum_{m=1}^{M}\gamma_m(t)\Big[\sum_i x_i\boldsymbol{A}_i\boldsymbol{o}(t)-\boldsymbol{\mu}_m\Big]^{\mathrm{T}}\\&\quad\boldsymbol{\Sigma}_m^{-1}\Big[\sum_i x_i\boldsymbol{A}_i\boldsymbol{o}(t)-\boldsymbol{\mu}_m\Big]\Big\}\end{aligned} \tag{6}$$

式中,T 表示语音段 s 中含有的总帧数,声学模型采用隐马尔可夫模型,共含有 M 个高斯混元,$\boldsymbol{\mu}_m$ 和 $\boldsymbol{\Sigma}_m$ 分别为第 m 个混元的均值矢量及协方差矩阵,$\gamma_m(t)$ 表示第 t 帧特征矢量属于第 m 个高斯混元的后验概率,可采用 Baum-Welch 前后向算法计算得到.

令似然度函数

$$\begin{aligned}Q(\boldsymbol{x})&=-\frac{1}{2}\sum_{t=1}^{T}\sum_{m=1}^{M}\gamma_m(t)\Big[\sum_i x_i\boldsymbol{A}_i\boldsymbol{o}(t)-\boldsymbol{\mu}_m\Big]^{\mathrm{T}}\\&\quad\boldsymbol{\Sigma}_m^{-1}\Big[\sum_i x_i\boldsymbol{A}_i\boldsymbol{o}(t)-\boldsymbol{\mu}_m\Big],\\\boldsymbol{\xi}_t&=[\boldsymbol{A}_1\boldsymbol{o}(t),\boldsymbol{A}_2\boldsymbol{o}(t),\cdots,\boldsymbol{A}_R\boldsymbol{o}(t)]\\&=[\boldsymbol{O}_1(t),\boldsymbol{O}_2(t),\cdots,\boldsymbol{O}_R(t)],\end{aligned}$$

则式(6)可转换为式(7):

$$\begin{aligned}\hat{\boldsymbol{x}}&=\arg\max_{x}\Big\{-\frac{1}{2}\sum_{t=1}^{T}\sum_{m=1}^{M}\gamma_m(t)[\boldsymbol{\xi}_t\boldsymbol{x}-\boldsymbol{\mu}_m]^{\mathrm{T}}\boldsymbol{\Sigma}_m^{-1}[\boldsymbol{\xi}_t\boldsymbol{x}-\boldsymbol{\mu}_m]\Big\}\\&=\arg\max_{x}\Big[-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{x}+\boldsymbol{f}^{\mathrm{T}}\boldsymbol{x}+C\Big]\end{aligned} \tag{7}$$

由式(7)可知,基于分段的区分性特征变换是一个典型的二次优化问题,其求解方法为:对式(7)中的似然函数关于 $\boldsymbol{x}$ 求导,并令导数等于0,C 是与变量 $\boldsymbol{x}$ 无关的常数项,可得式(8):

$$\hat{\boldsymbol{x}}=\boldsymbol{G}^{-1}\boldsymbol{f} \tag{8}$$

其中,

$$\begin{aligned}\boldsymbol{G}&=\sum_{m=1}^{M}\Big[\sum_{t=1}^{\mathrm{T}}\gamma_m(t)\boldsymbol{\xi}_t^{\mathrm{T}}\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\xi}_t\Big]\\&=\begin{bmatrix}g(1,1)&g(1,2)&\cdots&g(1,R)\\g(2,1)&g(2,2)&\cdots&g(2,R)\\\vdots&\vdots&\ddots&\vdots\\g(R,1)&g(R,2)&\cdots&g(R,R)\end{bmatrix}\end{aligned} \tag{9}$$

$$\begin{aligned}\boldsymbol{f}&=\sum_{m=1}^{M}\Big[\sum_{t=1}^{\mathrm{T}}\gamma_m(t)\boldsymbol{\xi}_t^{\mathrm{T}}\Big]\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu}_m\\&=[f(1)\quad f(2)\quad\cdots\quad f(R)]^{\mathrm{T}}\end{aligned} \tag{10}$$

式中,$g(i,j)=\sum_{m=1}^{M}\Big[\sum_{t=1}^{\mathrm{T}}\gamma_m(t)\boldsymbol{O}_i^{\mathrm{T}}(t)\boldsymbol{\Sigma}_m^{-1}\boldsymbol{O}_j(t)\Big]$,$f(k)=\sum_{m=1}^{M}\Big[\sum_{t=1}^{\mathrm{T}}\gamma_m(t)O_k^{\mathrm{T}}(t)\Big]\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu}_m$,$i,j,k=1,2,\cdots,R$. $g(i,j)$ 为第 i 个基矢量和第 j 个基矢量与观测数据的加权二阶统计量,$f(k)$ 为第 k 个基矢量与观测数据的加权一阶统计量.

2.2.2 联合变换矩阵和偏移矢量字典的特征变换

进一步,讨论字典中同时含有变换矩阵和偏移矢量的特征变换参数求解. 设经过域划分后变换矩阵 $\boldsymbol{A}_i$ 对应的权重系数为 x_i,共有 L 个偏移矢量,偏移矢量 $\boldsymbol{b}_j$ 所对应的权重系数为 y_j,则求解 $\hat{\boldsymbol{z}}=[\hat{\boldsymbol{x}},\hat{\boldsymbol{y}}]^{\mathrm{T}}$ 的问题可以转换为式(11)优化问题[14,15]:

$$\begin{aligned}\hat{\boldsymbol{z}}&=\arg\max_{x,y}\sum_{t=1}^{T}\sum_{m=1}^{M}\gamma_m(t)\log p(\boldsymbol{o}'(t)|\boldsymbol{\mu}_m)\\&=\arg\max_{x,y}\Big\{-\frac{1}{2}\sum_{t=1}^{T}\sum_{m=1}^{M}\gamma_m(t)\Big[\sum_i x_i\boldsymbol{A}_i\boldsymbol{o}(t)\\&\quad+\sum_j y_j\boldsymbol{b}_j-\boldsymbol{\mu}_m\Big]^{\mathrm{T}}\boldsymbol{\Sigma}_m^{-1}\\&\quad\Big[\sum_i x_i\boldsymbol{A}_i\boldsymbol{o}(t)+\sum_j y_j\boldsymbol{b}_j-\boldsymbol{\mu}_m\Big]\Big\}\end{aligned} \tag{11}$$

可令 $\boldsymbol{\xi}_{c,t}=[\boldsymbol{O}_1(t),\boldsymbol{O}_2(t),\cdots,\boldsymbol{O}_R(t),\boldsymbol{b}_1,\boldsymbol{b}_2,\cdots,\boldsymbol{b}_L]$,$\boldsymbol{z}=[x_1,x_2,\cdots,x_R,y_1,y_2,\cdots,y_L]$,则目标函数可以转换为式(12):

$$\begin{aligned}\hat{\boldsymbol{z}}&=\arg\max_{z}\Big\{-\frac{1}{2}\sum_{t=1}^{T}\sum_{m=1}^{M}\gamma_m(t)[\boldsymbol{\xi}_{c,t}\boldsymbol{z}-\boldsymbol{\mu}_m]^{\mathrm{T}}\cdot\\&\quad\boldsymbol{\Sigma}_m^{-1}[\boldsymbol{\xi}_{c,t}\boldsymbol{z}-\boldsymbol{\mu}_m]\Big\}\\&=\arg\max_{z}\Big[-\frac{1}{2}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{G}_c\boldsymbol{z}+\boldsymbol{f}_c^{\mathrm{T}}\boldsymbol{z}+C\Big]\end{aligned} \tag{12}$$

其中,

$$\boldsymbol{G}_c=\sum_{m=1}^{M}\Big[\sum_{t=1}^{\mathrm{T}}\gamma_m(t)\boldsymbol{\xi}_{c,t}^{\mathrm{T}}\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\xi}_{c,t}\Big],$$

$$\boldsymbol{f}_c=\sum_{m=1}^{M}\Big[\sum_{t=1}^{\mathrm{T}}\gamma_m(t)\boldsymbol{\xi}_{c,t}^{\mathrm{T}}\Big]\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu}_m,$$

可得到解的类似表达形式(13):

$$\hat{\boldsymbol{z}}=\boldsymbol{G}_c^{-1}\boldsymbol{f}_c \tag{13}$$

由于本文构造的字典具有一定的冗余性,在对每一语音段进行特征变换时,相比于未知数所拥有的数据量很有限. 在求解式(8)和(13)过程中,如何利用有限的数据从一个过完备的变换矩阵字典中,选取最佳的变换矩阵及其组合系数是本文的一个关键问题. 压缩感知中的匹配追踪算法能较好地解决该问题,接下来将结合匹配追踪算法求解目标函数. 由于式(8)和(13)求解过程相类似,下文中将主要介绍式(8)的求解过程,类似可以得到式(13)的解.

3 基于不相关匹配追踪算法的目标函数求解

借鉴正交匹配追踪(Orthogonal Matching Pursuit, OMP)[16,17]的算法思想,与最小化逼近误差作为目标函数不同,本文要使得似然度最大化,将似然度的变化率

定义为误差,同时采用字典项间的相关性代替正交性,得到一种不相关的匹配追踪算法. 这里字典项为变换矩阵 $\boldsymbol{A}_i$,其选取过程体现在特征 $\boldsymbol{o}(t)$ 经过矩阵 $\boldsymbol{A}_i$ 变换后的特征矢量 $\boldsymbol{O}_i(t)$ 上. 同样采用迭代的方式求解目标函数,每次迭代包含三个步骤:第一步从大字典中选取一个使得似然度提升量最大的字典项加入到支撑集中;第二步判断所选的字典项是否与支撑集中的字典项相关;第三步更新支撑集中字典项所对应的系数. 接着给出每一步骤的推导和求解过程.

3.1　最大似然字典项选取

支撑集选取过程为每次加入一个新字典项,所加入的字典项需使得似然度的增量值最大. 第一次选取时只需满足似然度最大即可,此时 $x_i = [g(i,i)]^{-1} f(i), i = 1,2,\cdots,K, K$ 为字典的大小. 将 x_i 代入目标函数 $Q(x)$ 中,得到特征经过第 i 个变换矩阵后的似然度式(14):

$$Q^1(x_i) = \frac{1}{2}[g(i,i)]^{-1} f(i)^2 + C \tag{14}$$

根据 $Q^1(x_i)$ 使之最大,确定第一个基矢量 $\boldsymbol{O}_{r_1}(t)$ 的序号 r_1 为式(15):

$$r_1 = \arg\max_i Q^1(x_i) \tag{15}$$

接着,每次在已选的支撑集中加入一个变换矩阵字典项,根据其权重系数进行加权组合特征变换,使得变换后的特征能获得最大的似然度提升. 假设第 k 次迭代后所得到的支撑集为 $D_k = \{\boldsymbol{O}_1(t), \boldsymbol{O}_2(t), \cdots, \boldsymbol{O}_k(t)\}$,其对应的加权系数为 $\boldsymbol{x}_k$,构成子空间 $\Gamma^k = \mathrm{span}\{\boldsymbol{O}_1(t), \boldsymbol{O}_2(t), \cdots, \boldsymbol{O}_k(t)\}$. 在字典 D 剩下的变换矩阵中进行第 $k+1$ 次迭代,选取字典项 $\boldsymbol{O}_l(t) \in D \backslash D_k$,其对应的系数为 x_l,此时似然度目标函数为式(16):

$$Q^{k+1}(x_l) = -\frac{1}{2}\sum_{t=1}^{T}\sum_{m=1}^{M}\gamma_m(t)[\boldsymbol{\xi}_t\boldsymbol{x}_k + x_l\boldsymbol{O}_l(t) - \boldsymbol{\mu}_m]^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_m^{-1}[\boldsymbol{\xi}_t\boldsymbol{x}_k + x_l\boldsymbol{O}_l(t) - \boldsymbol{\mu}_m] \tag{16}$$

令 $\dfrac{\partial Q^{k+1}(x_l)}{\partial x_l} = 0$,可以得到系数值 x_l 为式(17):

$$x_l = g(l,l)^{-1}\left[f(l) - \sum_{i=1}^{k} x_i g(l,i)\right] \tag{17}$$

将 $[\boldsymbol{x}_k, x_l]$ 代入似然度目标函数中,可得第 $k+1$ 次迭代后似然度的提升量 $\Delta Q^{k+1}(x_l)$:

$$\Delta Q^{k+1}(x_l) = Q^{k+1}(x_l) - Q^k(\boldsymbol{x}_k) = \frac{1}{2}g(l,l)^{-1}\left[f(l) - \sum_{i=1}^{k} x_i g(l,i)\right]^2 \tag{18}$$

其中,$Q^k(\boldsymbol{x}_k)$ 为第 k 次迭代后得到的似然度. 为使得似然度提升量最大,则第 $k+1$ 次所选择的字典项 $\boldsymbol{O}_{r_{(k+1)}}(t)$ 其相应的序号为式(19):

$$r_{(k+1)} = \arg\max_l \Delta Q^{k+1}(x_l) \tag{19}$$

3.2　相关基矢量的去除

由于本文的核心问题在于计算 $\hat{\boldsymbol{x}} = \boldsymbol{G}^{-1}\boldsymbol{f}$,为了保证解的稳定性以及加快收敛速度,特征变换矩阵需满足非奇异性,即矩阵 $\boldsymbol{G}$ 的列之间不具有相关性,这与常用的稀疏逼近中要求各组基之间正交化有所差别. 文献[18]通过引入一个不相关变换矩阵 $\boldsymbol{T}$ 来降低列之间的相关性,但由于本文变换矩阵字典项具有明确的物理意义,对变换矩阵进一步变换处理后较难保证其性能,另外,$\boldsymbol{T}$ 的求解过程需要计算逆矩阵,随着字典项的增大运算复杂度较高. 这里采用相关性度量来描述新选择的字典项与已有支撑集中字典项的相关性大小,如果新选择的字典项与已有字典项相关性过大,则需要将该字典项加入到冗余基集合 $\boldsymbol{\Phi}$ 中,不参与剩余的迭代.

假设新选择的特征矢量 $\boldsymbol{O}_l(t)$ 与支撑集中已有的特征矢量 $\{\boldsymbol{O}_i(t)\}_{i=1}^k$ 线性相关,则 $\boldsymbol{O}_l(t)$ 可以表示为:

$$\boldsymbol{O}_l(t) = \sum_{i=1}^{k}\alpha_i\boldsymbol{O}_i(t), \quad l = k+1, k+2, \cdots, K \tag{20}$$

其中,$\alpha_i \in R$,且 $\alpha_i, i = 1,\cdots,k$ 不是全 0. 同时由 $g(j,l)$ 的定义式可知,如果矢量 $\boldsymbol{O}_l(t)$ 能够被 $\{\boldsymbol{O}_i(t)\}_{i=1}^k$ 线性表示,则易得到 $g(j,l)$ 也能被 $\{g(j,i)\}_{i=1}^k$ 线性表示,$g(j,l) = \sum_{i=1}^{k}\alpha_i g(j,i), j = 1,2,\cdots,k$.

定义 $\boldsymbol{v}_l^k = [g(1,l) \quad g(2,l) \quad \cdots \quad g(k,l)]^{\mathrm{T}}$ 为新选择的矢量 $\boldsymbol{O}_l(t)$ 与支撑集中已有的矢量 $\{\boldsymbol{O}_i(t)\}_{i=1}^k$ 加权内积得到的列矢量. 由于 $\boldsymbol{v}_l^k$ 中的每一个元素 $g(k,l)$ 均能被 $\boldsymbol{G}_k$ 所对应列中的元素 $\{g(j,i)\}_{i=1}^k$ 线性表示,因此 $\boldsymbol{v}_l^k$ 与矩阵 $\boldsymbol{G}_k$ 中的列矢量线性相关,即 $\boldsymbol{v}_l^k = \boldsymbol{G}_k\boldsymbol{\alpha}, \boldsymbol{\alpha} = [\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_k]^{\mathrm{T}}$. 基于矢量 $\boldsymbol{O}_l(t)$ 根据式(9)可以计算得到 $g(l,l)$. 假设特征矢量 $\boldsymbol{O}_l(t)$ 与 $\{\boldsymbol{O}_i(t)\}_{i=1}^k$ 线性相关,根据上述分析过程,基于 $\{g(j,i)\}_{i=1}^k, g(l,l)$ 同时可以表示为 $\tilde{g}(l,l)$:

$$\tilde{g}(l,l) = \sum_{i=1}^{k}\sum_{j=1}^{k}\alpha_i\alpha_j g(i,j) = (\boldsymbol{v}_l^k)^{\mathrm{T}}\boldsymbol{\alpha} = (\boldsymbol{v}_l^k)^{\mathrm{T}}\boldsymbol{c}_l^k \tag{21}$$

式中,$\boldsymbol{\alpha} = \boldsymbol{G}_k^{-1}\boldsymbol{v}_l^k = \boldsymbol{c}_l^k$.

因此,可以先通过 $\boldsymbol{O}_l(t)$ 计算得到实际 $g(l,l)$ 的值,再利用其它基矢量进行线性表示得到其估计值 $\tilde{g}(l,l)$,两者之差用来表示新选择的矢量 $\boldsymbol{O}_l(t)$ 与支撑集中已有基矢量 $\{\boldsymbol{O}_i(t)\}_{i=1}^k$ 的相关性大小. $g(l,l) - (\boldsymbol{v}_l^k)^{\mathrm{T}}\boldsymbol{c}_l^k$ 值越大,则特征基矢量间的相关性越小;值越小,则相关性越大.

3.3　变换矩阵权重系数的更新

上一步骤中,如果所选择的特征基矢量满足不相关的要求,则将第 $k+1$ 次选择的 $\boldsymbol{O}_{r_{(k+1)}}(t)$ 加入支撑集中 $D_{k+1} = D_k \cup \{\boldsymbol{O}_{r_{(k+1)}}(t)\}$,此时 $D_{k+1} = \{\boldsymbol{O}_1(t), \boldsymbol{O}_2(t), \cdots,$

$\boldsymbol{O}_{k+1}(t)\}$,其中 $\boldsymbol{O}_{r_{(k+1)}}(t)$ 与 $\boldsymbol{O}_{k+1}(t)$ 是同一个基矢量,$r_{(k+1)}$ 为特征矢量在原大字典中的位置,而 $k+1$ 是指该基矢量在支撑集中的位置. 在加入新基矢量后,此时的变换矩阵系数 $[\boldsymbol{x}_k, x_{r_{(k+1)}}]^{\mathrm{T}}$ 并不是最优的. 根据匹配追踪算法的基本原理,需要在新的变换子空间 $\boldsymbol{I}^{k+1}=\mathrm{span}\{\boldsymbol{O}_1(t),\boldsymbol{O}_2(t),\cdots,\boldsymbol{O}_{k+1}(t)\}$ 中重新确定其权重系数. 依据式(8)确定其变换系数 $\boldsymbol{x}_{k+1}=\boldsymbol{G}_{k+1}^{-1}\boldsymbol{f}_{k+1}$. 根据 $\boldsymbol{v}_l^k$ 的定义,第 $k+1$ 次迭代新选择的矢量 $\boldsymbol{O}_{k+1}(t)$ 与支撑集中基矢量 $\{\boldsymbol{O}_i(t)\}_{i=1}^{k}$ 进行加权内积可以得到 $\boldsymbol{v}_{k+1}^k=[g(1,k+1)\quad g(2,k+1)\quad g(k,k+1)]^{\mathrm{T}}$. 根据矩阵 $\boldsymbol{G}_k$ 能进一步得到矩阵 $\boldsymbol{G}_{k+1}$ 的表达式:

$$\boldsymbol{G}_{k+1}=\begin{bmatrix}\boldsymbol{G}_k & \boldsymbol{v}_{k+1}^k\\ (\boldsymbol{v}_{k+1}^k)^{\mathrm{T}} & g(k+1,k+1)\end{bmatrix} \tag{22}$$

在确定新变换系数的过程中需要进行求逆运算,运算复杂度较高,特别是随着迭代次数和所加入基矢量数目的增加,这个问题会变得尤为突出. 幸运的是,$\boldsymbol{G}_{k+1}$ 可以由矩阵 $\boldsymbol{G}_k$ 和向量 $\boldsymbol{v}_{k+1}^k$ 有效地表示,根据分块矩阵求逆运算的相应理论,$\boldsymbol{G}_{k+1}$ 的逆矩阵 $\boldsymbol{G}_{k+1}^{-1}$ 可以利用第 k 步的结果 $\boldsymbol{G}_k^{-1}$ 快速地计算:

$$\boldsymbol{G}_{k+1}^{-1}=\begin{bmatrix}\boldsymbol{G}_k^{-1}+\beta_{k+1}^k\boldsymbol{c}_{k+1}^k(\boldsymbol{c}_{k+1}^k)^{\mathrm{T}} & -\beta_{k+1}^k\boldsymbol{c}_{k+1}^k\\ -\beta_{k+1}^k(\boldsymbol{c}_{k+1}^k)^{\mathrm{T}} & \beta_{k+1}^k\end{bmatrix} \tag{23}$$

其中,$\beta_{k+1}^k=[g(k+1,k+1)-(\boldsymbol{v}_{k+1}^k)^{\mathrm{T}}\boldsymbol{c}_{k+1}^k]^{-1}$,$\boldsymbol{c}_{k+1}^k=\boldsymbol{G}_k^{-1}\boldsymbol{v}_{k+1}^k$. 由 β_{k+1}^k 的定义式可知,若第 $k+1$ 次选择的基矢量与已有的字典项线性相关,则会使得 $g(k+1,k+1)-(\boldsymbol{v}_{k+1}^k)^{\mathrm{T}}\boldsymbol{c}_{k+1}^k$ 趋于零,而让 β_{k+1}^k 过大造成 $\boldsymbol{G}_{k+1}^{-1}$ 不稳定. 这也说明了在特征基矢量的选取过程中去除相关基矢量的必要性.

4 测试评估

4.1 实验设置

将本文分段区分性特征变换方法应用到连续语音识别中. 实验语料采用中文微软语料库 Speech Corpora (Version 1.0),其全部语料在安静办公室环境下录制,采样率为 16kHz,16bit 量化. 训练集共有 19688 句,共 454315 个音节,总时长约为 33 小时,测试集共 500 句,约为 0.7 小时,说话内容来自新闻报纸,对中文音节全覆盖. 文中选择声韵母作为模型基元,零声母(_a、_o、_e、_i、_u、_v),加上静音(sil)以及常规的声韵母,一共有 69 个模型基元,在此基础上将模型基元扩展为上下文相关的交叉词三音子(cross-word tri-phone). 基于 HTK 3.4.1 建立基线系统,声学模型采用 3 状态的隐马尔科夫模型,通过决策树对三音子模型进行状态绑定,绑定后的模型有效状态数为 2843 个. 利用 SRILM 工具根据语料库中自有的标注文件训练得到语言模型. 文中均采用有调音节的识别准确率进行识别性能的评估.

4.2 基于帧特征变换方法的识别性能

这里采用 13 维的 MFCC 特征联合当前帧及其前后各 4 帧共 9 帧,并采用 MLLT + LDA 作为初始的变换矩阵,进行最大似然声学模型的建立. 特征变换中全局 GMM 模型是由声学模型状态中的高斯聚类得到,最终共有 800 个高斯. 在此基础上,分别得到了基于词图信息和基于状态绑定的 fMPE、m-fMPE、RDLT 特征变换方法的识别性能,并进一步讨论了当声学模型分别采用最大似然和区分性训练(Boosted Maximum Mutual Information,BMMI)时,各种特征变换方法的识别性能,具体识别结果表 1 所示.

表 1 不同特征变换方法的识别准确率(%)

特征变换方法	LDA + MLLT	fMPE		m-fMPE		RDLT	
	lattice	lattice	tied-state	lattice	tied-state	lattice	tied-state
ML	74.28	75.91	75.83	76.34	76.56	76.67	76.92
BMMI	76.52	77.04	77.69	77.47	78.21	77.85	78.66

由表 1 中的识别结果可知,区分性特征变换方法的识别性能均较为明显地优于线性判别分析方法. 基于词图信息和状态绑定的 fMPE 方法得到的识别结果相当. 为了保证 fMPE 的性能其所需的高斯混元数为 12000 个,所得到的特征变换矩阵为其他方法的 15 倍左右,这主要是因为其每一个域中所含有的参数和信息量较小,需要增大域的个数以保证信息量. 由于它利用前后相关的后验概率信息进行特征变换,采用状态绑定的方式会在一定程度上影响这种前后相关性的获取. m-fMPE,RDLT 采用状态绑定的方式得到识别结果会优于采用词图信息的方式. 在特征变换的基础上,对声学模型区分性训练后识别性能得到进一步提升,且基于状态绑定的特征变换方法其优势更为明显. 这说明采用状态绑定方法进行特征变换时,可以有效地克服声学模型对特征变换的影响,在求解优化过程中侧重于寻找区分性特征.

4.3 基于域划分变换矩阵字典项的识别性能

首先基于变换矩阵 $\boldsymbol{A}$ 构造字典,字典共有 800 个字典项,采用不相关匹配追踪算法进行特征变换. 在这个过程中,语音信号的分段时长、匹配追踪算法中的似然度增量阈值 δ 直接决定着变换矩阵的选取,进而影响识别性能,因此分别讨论了上述参数在不同设置条件下的识别性能,所选字典数的上限 $N=200$. 通常语音分段以帧级单元为基础,通过某种分段方式来构造,常用的分段方式有两种:一是固定长度分段,即按照指定的长度进行分割;二是自适应长度分段,即对语音信号按照某种关联准则进行划分,例如,采用强制对齐的方式进行分段,这种分段考虑了语音特征空间内在的关联

关系,是常用的分段对齐方法. 这里将测试集强制对齐到前 800 个状态中进行分段,分段后语音分段时长均值为 3.15s,方差为 1.47,接着分别讨论了两种分段方式的识别性能. 表 2 给出了不同分段时长、似然度增量阈值条件下,RDLT 变换的连续语音识别率,其中加黑字体为除强制对齐外最好的识别结果,括号内为稀疏度,其度量方式为零系数占所有系数的比例.

表 2　不同分段时长、似然度增量阈值的识别准确率及其稀疏度(%)

字典项	参数设置	分段时长					
		1s	2s	2.5s	3s	4s	强制对齐
A	$\delta=10$	74.68 (88.57)	**76.51** (**87.34**)	76.03 (84.19)	75.11 (81.08)	73.89 (75.89)	76.98 (85.96)
	$\delta=20$	75.39 (91.02)	**77.09** (**89.43**)	76.94 (86.59)	75.67 (83.31)	74.53 (78.07)	77.41 (88.61)
	$\delta=30$	76.09 (92.21)	**77.22** (**90.68**)	77.10 (87.44)	76.38 (84.56)	75.13 (79.11)	77.70 (89.67)
	$\delta=40$	76.46 (94.18)	**77.45** (**91.57**)	77.21 (88.32)	76.59 (85.17)	75.64 (81.06)	78.04 (90.35)
	$\delta=50$	76.18 (95.27)	**77.17** (**92.93**)	77.04 (89.47)	76.15 (86.45)	75.42 (82.25)	77.82 (91.76)

由表 2 的识别结果可知,在相同的似然度增量阈值 δ 下,当数据分段较小时,稀疏度较大;随着数据量的增大,所选择的特征变换矩阵也会增多;在相同的数据分段长度下,δ 值越大,稀疏度越大,所选择的变换矩阵越少,此时主要选择对识别性能影响较大的特征变换矩阵,这与前面的分析是相符的. 稀疏度过大和过小都不能获得最佳的识别性能. 同时由表中的识别结果可知,对语音采用不同的分段方式识别结果会有较大的差异. 在对语音采用固定长度的分段方法中,将语音分成 2s 一段的方式能得到最优的识别结果,这主要是采用这一长度能得到相对稳定的统计特性,得到的参数信息较为准确. 随着分段长度的增大,识别性能反而会开始下降,这主要是因为当数据分段过大时段内的声学性质会有较大的差异,即使是数据较为充分也难以获得较好的参数估计,来同时描述差异性较大的语音信号段,此时应该将语音信号段进一步细分,分别估计变换矩阵. 采用强制对齐的分段方法能得到最高的识别性能,这主要是因为对齐到相同状态的数据具有相类似的声学特性,利用这些数据能估计得到稳健的参数信息.

采用匹配追踪算法还能根据所拥有的数据量大小,自适应地确定变换基矢量的数量,有效地避免常用方法中需要对基矢量个数进行经验设定. 由于本文是一个凸优化问题,初值的设置对识别结果的影响不大. 匹配追踪算法具有很高的运算效率,这很适合于前端的特征变换,不会给识别系统中引入太多的耗时,减小对后端识别解码的影响. 由于识别算法是一个非线性过程,较难直接得到其理论的计算复杂度,通过分别定性地统计特征变换和整个识别算法的耗时,得知特征变换的耗时占整个识别算法耗时的 1% 以下,对整个识别算法的影响不大.

4.4　联合变换矩阵和偏移矢量字典项的识别性能

由表 1 的实验结果可知,m-fMPE 和 RDLT 能得到相对较优的识别性能,m-fMPE 侧重于偏移矢量的求解,而 RDLT 能得到更好的变换矩阵,两者具有一定的互补性. 由于匹配追踪算法具有较高的运算效率,接下来将两者变换矩阵结合起来,构造一个过完备字典,字典共有 1600 个字典项,采用强制对齐的方式进行数据的分段,利用不相关匹配追踪算法进行变换矩阵的选取及其系数的确定,实验结果表 3 所示,其中 $\boldsymbol{A}$ 是 RDLT 方法得到的变换矩阵,$\boldsymbol{M}$ 是 m-fMPE 方法得到的变换矩阵,$\boldsymbol{b}$ 是对矩阵 $\boldsymbol{M}$ 进行分解后对应的偏移矢量,括号内为稀疏度.

表 3　联合不同变换矩阵和偏移矢量字典项的识别准确率及其稀疏度(%)

字典项	$\boldsymbol{A}$	$\boldsymbol{M}$	$\boldsymbol{A}+\boldsymbol{M}$	$\boldsymbol{A}+\boldsymbol{b}$
ML	78.04 (90.35)	77.57 (75.46)	78.16 (85.87)	78.55 (87.38)
BMMI	80.18	79.15	80.43	80.89

由表 3 的识别结果可知,当只采用一组字典时,采用变换矩阵 $\boldsymbol{A}$ 能得到最好的性能,主要是由于变换矩阵 $\boldsymbol{A}$ 是矩阵 $\boldsymbol{M}$ 的一般化,其具有更强的描述能力,这同时说明在进行特征变换时,变换矩阵比偏移矢量能

更好地保证性能. 结合变换矩阵和偏移矢量构成过完备字典进行特征变换,其得到的性能会优于仅采用一组字典的方法,表明这两组字典具有一定的互补性,选择的变换矩阵和偏移矢量个数介于采用单组字典 $\boldsymbol{A}$ 和 $\boldsymbol{M}$ 之间. $\boldsymbol{A}+\boldsymbol{b}$ 的方法会好于 $\boldsymbol{A}+\boldsymbol{M}$ 的方法,这主要是由于 $\boldsymbol{M}$ 矩阵中也含有变换矩阵,这与 $\boldsymbol{A}$ 中的变换矩阵会存在部分重复,而使得这部分变换矩阵的权值过大,造成过分重视,降低识别性能. 仅利用其偏移矢量 $\boldsymbol{b}$ 结合 $\boldsymbol{A}$ 构造字典,能获得最高的识别性能. 在特征变换的基础上,对声学模型区分性训练均能进一步提高识别性能.

5 结论

本文提出了一种基于语音分段的区分性特征变换方法,在特征变换求解过程中,引入了压缩感知中的稀疏逼近相关理论. 通过采用状态绑定的方式,求解变换矩阵和偏移矢量构造过完备的字典. 根据不相关匹配追踪算法,将特征变换的似然度作为目标函数,在目标函数的优化过程中选择最佳的特征变换矩阵及其组合系数. 实验结果表明,相比于传统基于帧的特征变换方法,本文方法能够有效地提高识别性能,通过采用强制对齐的方式进行语音分段能得到最好的识别性能. 在特征变换的基础上,进行声学模型的区分性训练能进一步提升识别性能. 后续的研究可以将本文方法应用于其它特征变换方法中.

参考文献

[1] Abbasian H, Nasersharif B, Akbari A, et al. Optimized linear discriminant analysis for extracting robust speech features[A]. Proceedings of International Symposium Communication Control and Signal Processing[C]. Julians, Malta: IEEE, 2008. 819 – 824.

[2] Nasersharif B, Akbari A. SNR-dependent compression of enhanced Mel sub-band energies for compensation of noise effects on MFCC features[J]. Pattern Recognition Letters, 2011, 28 (11), 1320 – 1326.

[3] Povey D, Kingsbury B, Mangu L, et al. fMPE: Discriminatively trained features for speech recognition[A]. Proceedings of the International Conference on Audio, Speech and Signal Processing[C]. Philadelphia, United States: IEEE, 2005. 961 – 964.

[4] Zhang B, Matsoukas S, Schwartz R. Discriminatively trained region dependent feature transforms for speech recognition[A]. Proceedings of the International Conference on Audio, Speech and Signal Processing[C]. Toulouse, France: IEEE, 2006. 313 – 316.

[5] Zhang B, Matsoukas S, Schwartz R. Recent progress on the discriminative region-dependent transform for speech feature extraction[A]. Proceedings of the Annual Conference of International Speech Communication Association[C]. Pittsburgh, United States: ISCA, 2006. 1495 – 1498.

[6] Takashi F, Osamu I, Masafumi N, et al. Regularized feature-space discriminative adaptation for robust ASR[A]. Proceedings of the Annual Conference of International Speech Communication Association[C]. Singapore: ISCA, 2014. 2185 – 2188.

[7] Povey D. Improvements to fMPE for discriminative training of features[A]. Proceedings of the Annual Conference of International Speech Communication Association[C]. Lisbon, Portugal: ISCA, 2005. 2977 – 2980.

[8] Yan Z J, Huo Q, Xu J, et al. Tied-state based discriminative training of context-expanded region-dependent feature transforms for LVCSR[A]. Proceedings of the International Conference on Audio, Speech and Signal Processing[C]. Vancouver, Canada: IEEE, 2013. 6940 – 6944.

[9] Deng L, Chen J S. Sequence classification using the high-level features extracted from deep neural networks[A]. Proceedings of the International Conference on Audio, Speech and Signal Processing[C]. Florence, Italy: IEEE, 2014. 6894 – 6898.

[10] Ling Z H, Kang S Y, Zen H, et al. Deep learning for acoustic modeling in parametric speech generation: a systematic review of existing techniques and future trends[J]. IEEE Signal Processing Magazine, 2015, 32(3): 35 – 52.

[11] George S, Brian K. Discriminative feature-space transforms using deep neural networks[A]. Proceedings of the Annual Conference of International Speech Communication Association[C]. Oregon, United States: ISCA, 2012.

[12] Paulik M. Lattice-based training of bottleneck feature extraction neural networks[A]. Proceedings of the Annual Conference of International Speech Communication Association[C]. Lyon, France: ISCA, 2013. 89 – 93.

[13] Liu D Y, Wei S, Guo W, et al. Lattice based optimization of bottleneck feature extractor with linear transformation[A]. Proceedings of the International Conference on Audio, Speech and Signal Processing[C]. Florence, Italy: IEEE, 2014. 5617 – 5621.

[14] Kuhn R, Junqua J C, Nguyen P, et al. Rapid speaker adaptation in eigenvoice space[J]. IEEE Transactions on Speech and Audio Processing, 2000, 8(6): 695 – 707.

[15] Ghoshal A, Povey D, Agarwal M, et al. A novel estimation of feature-space MLLR for full-covariance models[A]. Proceedings of International Conference on Acoustics, Speech and Signal Processing[C]. Texas, USA: IEEE, 2010. 4310 – 4313.

[16] Mallat S G, Zhang Z. Matching pursuits with time-frequency dictionaries[J]. IEEE Transactions on Signal Processing, 1993, 41(12): 3397 – 3415.

[17] Tropp J A, Gilbert A C. Signal recovery from random measurement via orthogonal matching pursuit[J]. IEEE Transactions on Information Theory, 2007, 53(12): 4655 – 4666.

[18] Needell D, Vershynin R. Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit[J]. IEEE Journal of Selected Topics Signal Processing, 2009, 4(2): 310 – 316.

作者简介

陈　斌　男,1987 年生于江西萍乡. 现为解放军信息工程大学信息系统工程学院博士研究生,西南电子电信技术研究所上海分所工程师. 主要研究方向为连续语音识别、区分性训练和机器学习.

E-mail: chenbin873335@163.com

牛　铜　男,1982 年生于河南郑州. 现为解放军信息工程大学信息系统工程学院博士研究生. 主要研究方向为语音识别和语音增强.

E-mail: niutong0072@gmail.com