

# 基于瓶颈特征的藏语拉萨话连续语音识别研究

周楠<sup>1</sup> 赵悦<sup>1,†</sup> 李要墙<sup>1</sup> 徐晓娜<sup>1</sup> 才旺拉姆<sup>1</sup> 吴立成<sup>1</sup>

1. 中央民族大学 信息工程学院, 北京 100081;

† 通讯作者, E-mail: zhaoyueso@muc.edu.cn

**摘要** 目前神经网络在语音识别中得到了广泛的应用, 虽然其具有较高的稳健性和语义区分性, 但其后验特征无法用于成熟的高斯混合-隐马尔科夫声学建模框架中。然而, 具有狭窄瓶颈的神经网络可以解决此问题, 其提取的瓶颈特征不仅具有语音长时相关性和紧凑表示的特点, 并且可以代替传统的 MFCC 特征进行 GMM-HMM 声学建模。本文研究将瓶颈特征及其与 MFCC 的复合特征用于藏语连续语音识别任务中。实验表明, 瓶颈复合特征取得了较神经网络后验特征和单瓶颈特征更好的识别表现。

**关键词** 藏语拉萨话; 连续语音识别; 高斯混合-隐马尔科夫模型; 瓶颈特征; 神经网络

中图分类号 TP391

## Study on Continuous Speech Recognition based on Bottleneck Features for Lhasa-Tibetan Dialect

ZHOU Nan<sup>1</sup>, ZHAO Yue<sup>1,†</sup>, LI Yaoqiang<sup>1</sup>, XU Xiaona<sup>1</sup>, CAIWANG Lamu<sup>1</sup>, WU Licheng<sup>1</sup>

1. School of Information Engineering, Minzu University of China, Beijing, 100081, China;

†Corresponding Author, E-mail: zhaoyueso@muc.edu.cn

**Abstract** At present, deep neural network has been widely used in speech recognition, although it has high robustness and semantic distinction, but its posterior features cannot be used for GMM-HMM acoustic modeling framework. However, the neural network with a narrow bottleneck can solve this problem, and its bottleneck features not only have long term context-dependence and compact representation of speech signal, but also can replace the traditional MFCC features for GMM-HMM acoustic modeling. In this paper, we study on applying bottleneck features and its concatenated features with MFCC into Lhasa-Tibetan continuous speech recognition. The experimental results show that the concatenated features of bottleneck features and MFCC achieved better performance than the posterior features of deep neural network and mono-bottleneck features.

**Key words** Lhasa-Tibetan; continuous speech recognition; GMM-HMM; bottleneck features; DNN

神经网络是一个拥有很多隐层的机器学习模型, 该模型提取的深度特征具有分布平稳、表征能力强、易于建模等优点。目前深度学习理论已成功应用于大词汇量连续语音识别 (Large Vocabulary Continue Speech Recognition, 简称 LVCSR) 中<sup>[1-2]</sup>。尽管深度学习在英文、中文等主要语种的语音识别任务上的优势得到了实验验证, 但深度学习在藏语大词汇量连续语音识别任务中的应用尚未深入研究。文献<sup>[3-4]</sup>探讨了基于深度学习的藏语音素和藏语孤立词识别。文献<sup>[5]</sup>研究了基于神经网络 (Deep Neural Network, 简称 DNN) 的藏语拉萨话连续语音识别, 其使用声韵母作为识别基元, 利用 DNN 输出层特性进行三音素 HMM 声学建模, 识别模型单音节识别率达到 43% 左右。

在 DNN-HMM 声学模型应用之前, 传统的语音识别声学模型采用 MFCC 特征对 GMM-HMM 进行建模, GMM-HMM 模型具有完善的理论知识体系, 训练效率高, 但是由于每帧 MFCC 特征通常只包含毫秒

---

教育部人文社科规划基金项目 (15YJAZH120) 资助, 中央民族大学双一流学科建设项目资助

收稿日期: 2017-06-01; 修回日期: 2017-08-28; 网络出版时间:

网络出版地址:

的语音信号，信息量不足，容易受到噪声污染，其抗噪能力很弱<sup>[6]</sup>。

为了利用 GMM-HMM 的性能优势，文献<sup>[7-9]</sup>研究了一种具有狭窄中间层的 BottleNeck (BN) 深度神经网络，其提取了网络中间的瓶颈特征，可以替代传统的 MFCC 语音特征，用于训练传统的 GMM-HMM 模型，该特征不仅具有语音长时相关性，而且具有高度抽象表征信号的能力。基于该思想，本文研究基于瓶颈特征及其与 MFCC 复合特征的藏语连续语音识别技术，以验证瓶颈特征在藏语大词汇量连续语音识别工作中的有效性。

本文第一部分介绍了基于瓶颈神经网络的特征提取方法；第二部分介绍了藏语知识和藏语识别基元选取；第三部分介绍了实验过程、结果与分析；最后对本文进行了总结。

## 1 瓶颈特征提取

在语音信号处理过程中，传统的声学特征参数容易受到环境噪声、说话人差异性、信道的影响，即使相同的语音内容，在不同人不同发音时，都会有很大差异。相对底层声学参数，高层信息具有更强的鲁棒性。具有狭窄中间层的 BottleNeck 神经网络是一种特殊结构的 DNN，其对输入特征进行多次非线性变换，得到区分性更强的声学特征，模型中间层的 BN 特征可以替代传统的语音特征，进行 GMM-HMM 声学模型建模。

提取 BN 特征需要先训练一个 DNN 神经网络，第一步运用对比散度算法来训练受限玻尔兹曼机 (Restricted Boltzman Machine, 简称 RBM)，得到其参数；第二步将第一层输出作为第二层的输入训练第二层 RBM，得到第二层的参数，重复第二步，直达到所需深度。最后采用有监督的训练算法，利用误差反向传播算法来优化参数<sup>[10]</sup>。BN 层的结点比其他隐层少，在 DNN 训练完成后，将 BottleNeck 层的网络参数取出作为后续建模的瓶颈特征<sup>[11]</sup>。整个过程如图 1 所示。

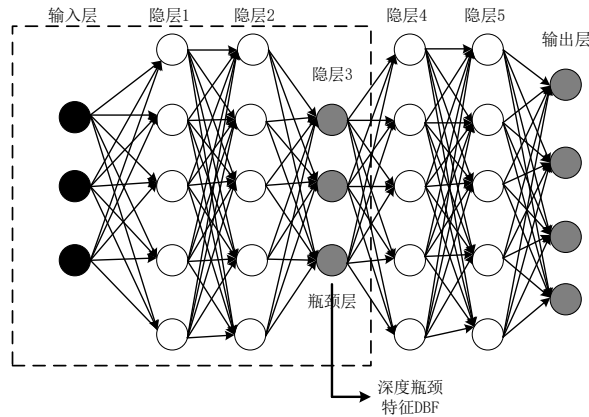


图 1 瓶颈结构深度神经组成结构图

Fig.1 BN-DNN Composition structure diagram

### 1.1 RBM 实现深度学习

首先定义受限玻尔兹曼机 (Restricted Boltzman Machine, 简称 RBM)，对于变量  $v, h$ ， $v = [v^1, v^2, \dots, v^n]^T$ ， $h = [h^1, h^2, \dots, h^m]^T$ ， $v^i, h^j \in \{0, 1\}$ ， $i = 1, 2, \dots, n$ ， $j = 1, 2, \dots, m$ ，如果其分布函数为

$$p(v, h) = \frac{1}{Z} e^{-E(v, h)}, \quad (1)$$

$$Z = \sum_{v, h} e^{-E(v, h)}, \quad (2)$$

RBM 的能量函数  $E(v, h)$  表示为：

$$E(v, h, \theta) = -\sum_{i=1}^n \sum_{j=1}^m w_{ij} v_i h_j - \sum_{i=1}^n a_i v_i - \sum_{j=1}^m b_j h_j, \quad (3)$$

则符合上式分布的一组随机变量称为受限玻尔兹曼机。从上式可以看出，可见结点的一组取值和隐藏结点的一组取值发生的概率  $p(v, h)$  是跟能量函数有关系的。可以得到条件概率如下：

$$p(v) = \frac{\sum_h e^{-E(v, h)}}{Z}, \quad (4)$$

$$p(h) = \frac{\sum_v e^{-E(v, h)}}{Z}, \quad (5)$$

$$p(v | h) = \frac{e^{-E(v, h)}}{\sum_v e^{-E(v, h)}}, \quad (6)$$

$$p(h | v) = \frac{e^{-E(v, h)}}{\sum_h e^{-E(v, h)}}, \quad (7)$$

由此引入一个自由能量概念可以得到最大似然值。自由能量函数：

$$FreeEnergy(v) = -\ln \sum_h e^{-E(v, h)}, \quad (8)$$

则：

$$p(v) = \frac{e^{-FreeEnergy(v)}}{Z}, \quad (9)$$

对上式两边进行取对数之后可得到：

$$\ln p(v) = -FreeEnergy(v) - \ln Z, \quad (10)$$

由前面的运算我们可以得到：

$$\sum_v \ln p(v) = \ln \left[ \prod_v p(v) \right], \quad (11)$$

其中  $\ln \left[ \prod_v p(v) \right]$  就是似然函数，最大似然估计就是使  $\ln \left[ \prod_v p(v) \right]$  最大，对于 RBM 系统，想要得到最大

似然值，就是求其所对应的参数  $\theta \in \{w_{ij}, a_i, b_j; i = 1, 2, \dots, n, j = 1, 2, \dots, m\}$  值，可以将 RBM 系统似然函数  $\prod_v p(v)$  写成公式 (12)，

$$L(\theta | v) = \prod_v p(v), \quad (12)$$

对上式进行取对数并求偏导数后得到：

$$\Delta w_{ij} = \varepsilon (\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model}), \quad (13)$$

公式 (13) 中， $\langle v_i h_j \rangle_{data}$  表示输入样本的自由能量期望值， $\langle v_i h_j \rangle_{model}$  表示模型产生的样本数据的自由能量期望值。

为了能够让系统似然函数达到最大，可通过上述训练找出一个能让系统的自由能量的总和最小的参数。这样通过 RBM 训练，才能更充分训练多层神经网络模型，从而得到一个最优模型<sup>[12]</sup>。

## 1.2 深度瓶颈特征提取

在完成 DNN 的训练之后，将瓶颈层以上的隐含层和输出层网络参数移除掉，得到深度瓶颈特征的提取函数  $g(*)$ ，其提取形式可表示为：

$$y = g(x), \quad (14)$$

其中  $y \in R^{D'} = [y_1, \dots, y_{D'}]^T$ ， $D'$  表示 BottleNeck 层的结点数，输入特征矢量  $x = [x_1, \dots, x_D]^T$ 。

$$y_{d'} = g(x|\theta) = \sum_j w_{d',j}^l f(\sum_i w_{j,i}^{l-1} f(\dots + f(\sum_d w_{d,i}^1 + b_i^1)\dots) + b_j^{l-1}) + b_{d'}^l, \quad (15)$$

$d' = 1, \dots, D'$  中的  $l$  表示 BottleNeck 的层数,  $w_{i,j}^l$  表示第  $l$  层的结点  $i$  与第  $l-1$  层的结点  $j$  之间的网络权重参数,  $b_i^l$  表示第  $l$  层的结点  $i$  的偏置,  $f(*)$  表示激活函数。

### 1.3 瓶颈复合特征

声学复合特征指的是将非短时差异特征与传统短时特征拼接过程。文献<sup>[13]</sup>将短时声学特征 MFCC 与多层感知器 (MLP) 产生的两类差异性特征拼接为新的特征参数, 利用新特征流对 GMM-HMM 进行声学建模, 在中文语音识别率上较单一特征有较大提升。然而多层感知器属于浅层神经网络, 不能提取深层的语音特征, 本文采用 DNN 提取具有长时性的 39 维瓶颈特征与传统的 39 维 MFCC 特征复合成为一个 78 维的高维特征参数, 并通过线性区分分析 (LDA)<sup>[14]</sup> 进行降维处理, 将降维后的 39 维特征参数用于 GMM-HMM 声学建模。其特征流复合过程如图 2 所示。

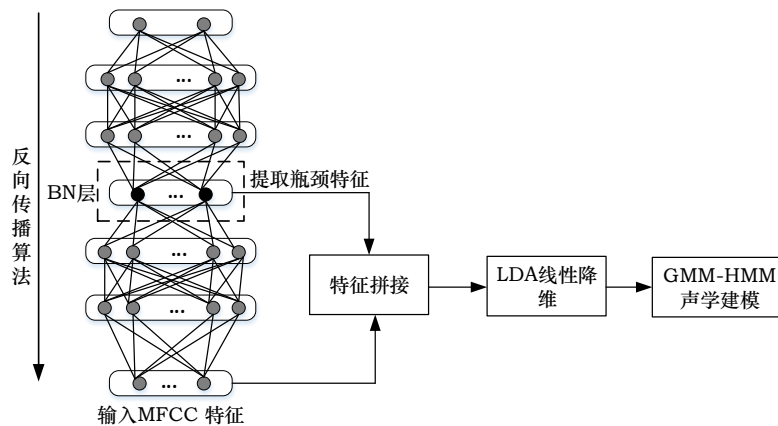


图 2 瓶颈复合特征提取

Fig.2 Extraction of concatenated bottleneck feature

## 2 藏语拉萨话音素集合

藏语隶属于汉藏语系藏缅语族支, 主要分布在中国西藏自治区、青海、甘肃、四川、云南等省份。全国大约有 600 多万人使用藏语。此外, 中国的邻国印度、巴基斯坦、尼泊尔等国家也有藏语分布。藏语有三种方言如: 卫藏、康巴和安多, 这三种方言虽然书面语相同, 但发音不同<sup>[15]</sup>。

藏语属于拼音文字, 音节也是藏语的基本单位, 由一个或几个音素按一定规律组合而成, 音节之间用音节符 (“.”) 分隔。其发音特点有自己的独特规律, 藏语共有 30 个辅音字母和 5 个元音符合, 辅音字母可以根据发音部分和发音方法来分, 而这些字母发音按照音势的强弱, 气流的多寡, 带音不带音以及发音肌肉的松紧来划分<sup>[16]</sup>。

藏文具有英文一样的书写规则, 整体自左向右横向书写, 但藏文一个音节内有竖向叠加的书写, 通常称为“叠加书写”, 叠加书写的结构中以一个辅音字母的中心位置, 将字母分为“前加字”, “上加字”, “下加字”, “后加字”, “再后加字”, 这个中心位置的辅音字母称之为“基字”, 如图 3 所示。

一个藏文音节中基字发的是主音, “前加字”, “上加字”, “下加字”, “后加字”, “再后加字” 发的是辅助音<sup>[17]</sup>。

由于藏语词汇量大, 常用的就有上万条, 并且存在各种变体。符合“字性组合法”的常用藏语单音节有 5600 多个, 所以, 词和单音节不合作为大词汇量连续语音识别的识别单元。

音素是组成音节的最小的语音片段, 藏语最初是一种按一字一音准确拼写的原则创始的拼音文字, 因此理论上藏语语音识别系统采用音素基元比较科学<sup>[15]</sup>。

本文整理的藏语拉萨话音素集包含 59 个音素, 并制定了相应的拉丁转写方案。如表 1 所示。

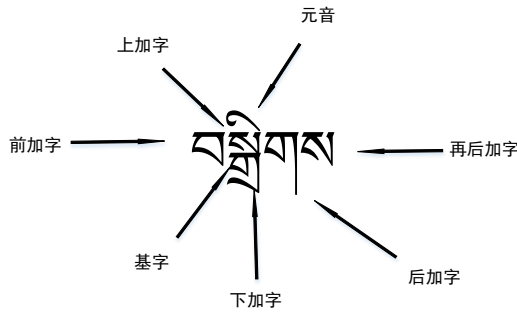


图3 藏文的音节结构图

Fig.3 The syllable structure of Tibetan language

表1 藏语拉萨话音素及拉丁转写表

Table 1 Lhasa-Tibetan dialect phonemes and Latin Transliteration table

IPA	拉丁转写	IPA	拉丁转写	IPA	拉丁转写	IPA	拉丁转写	IPA	拉丁转写
c	gy	nc	jy	c <sup>h</sup>	ky	h	h	a:	v
ʂ	hr	j	y	k	k	ŋk	gh	k <sup>h</sup>	g
l	l	l̥	lh	m	m	n	n	ŋ	ng
ŋ	ny	p	p	mp	bh	p <sup>h</sup>	b	r	r
t	t	nt	dh	ts <sup>h</sup>	tsh	tʂ	ts	nts <sup>h</sup>	th
tʂ <sup>h</sup>	chr	t <sup>h</sup>	d	te	c	nte	jh	te <sup>h</sup>	qj
br	br	ntʂ	z	w	w	ɛ	sh	f	ff
a	a	ʔ	ab	e	e	e <sup>ʔ</sup>	eb	e:	ew
ẽ	eu	ɛ	el	ɛ <sup>ʔ</sup>	elb	ɛ:	elw	Ø <sup>ʔ</sup>	fb
Ø	f	Ø:	fw	i	i	i <sup>ʔ</sup>	ib	ĩ	il
i:	jw	o	o	o:	ow	u	u	u:	uw
y:	yw	y <sup>ʔ</sup>	yb	ỹ	yu	s	s		

### 3 实验与结果分析

前文研究了深度神经网络瓶颈特征提取的工作原理，并给出了瓶颈复合特征提取的方法，结合建立的音素基元，本节将进行藏语拉萨话连续语音识别实验方案设计、性能验证和结果分析。

#### 3.1 语料库与评价指标

本文的语料库收集自藏语拉萨话发音标准的8个女生，8个男生的日常口语语音。选用的句子每句长度约为2~60秒，形成约10h的藏语连续语音语料库。录制过程中，采用16KHz采样频率、16bit量化精度。语料中的8.5h用作模型训练，1.5h做测试。

实验评价指标为藏语连续语音识别中的音节错误率（WER）。设 $N$ 为测试文本中词的数量， $I$ 为插入词个数， $D$ 为删除词个数， $S$ 为替代词个数。WER定义如式（16）所示：

$$WER = \frac{I + D + S}{N}, \quad (16)$$

#### 3.2 语言模型

由于语言模型在整个语音识别过程中的作用非常重要，有无语言模型对整个语音识别系统的使用范围和识别效率影响很大。目前，大部分大词汇量连续语音识别系统均采用基于n-gram的统计语言模型用于解码器中计算语言模型概率<sup>[18]</sup>。本文使用训练集中的文本来训练词的三元语言模型，并采用回退平滑技术来

解决数据稀疏的问题。

在搭建藏语语音识别系统时，我们使用开源工具箱 `kaldi`<sup>[19]</sup>进行相关实验，包括数据准备、声学特征提取、声学模型的训练与解码；使用 `DNN` 工具包进行相关 `DNN` 的搭建与训练。

### 3.3 GMM-HMM 模型

本文实验使用了三音素建模。对普通的三音素单元使用自左向右的无状态间跨越的三状态 `HMM`，而对静音模型，则采用状态间可跨越的五状态 `HMM`，每个 `HMM` 拓扑结构前后都有一个开始状态和一个结束状态。结合先前藏语拉萨话语音发音特征研究成果<sup>[20]</sup>，我们对拉萨话 59 个音素划分了 20 个类别集，并分别建立决策树问题集。

`GMM-HMM` 模型用最大似然估计准则来训练，输入是 39 维特征，帧长为 25ms，帧移为 10ms。`HMM` 中每个状态设置 100 个独立的高斯分量。`GMM-HMM` 模型结构如图 4 所示。

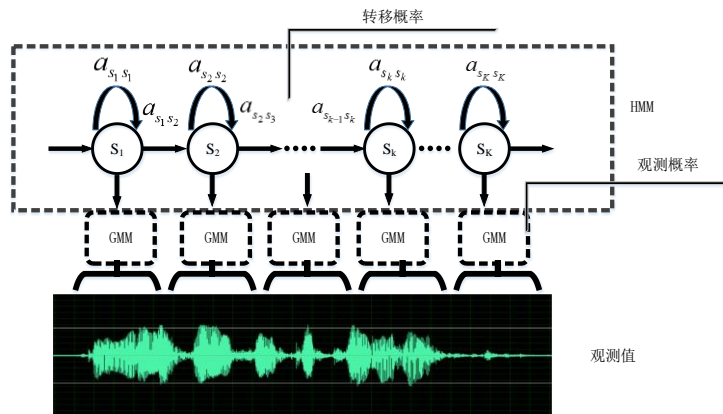


图 4 高斯混合-隐马尔科夫声学模型

Fig.4 GMM-HMM Model

### 3.4 BN-GMM-HMM 模型

传统的瓶颈特征是由多层感知器（`Multi-layer perception`, `MLP`）获取，它采用一种非线性的特征变换与降维技术<sup>[21]</sup>。本文基于 `DNN` 模型来生瓶颈声学特征，`DNN` 的训练特征使用 39 维的 `MFCC` 特征（12 维滤波器输出值加上 1 维对数能量，以及其一阶差分和二阶差分），并进行前后各 5 帧拼接，共 11 帧的上下文扩展。`BN-DNN` 网络总共 7 层，包含 1 个输入层，5 个隐含层和 1 个输出层，`BN` 层定义在中间一个隐含层。输入层包含 429 个结点，输出层包含 1153 个结点，每个结点对应不同的三音素，`BN` 层结点数和 `MFCC` 特征维数一样有 39 个结点，其它隐含层分别包含 1024 个结点。因此，`BN-DNN` 的结构为“429-1024-1024-39-1024-1024-1153”。对每个隐含层进行 100 次的 `RBM` 预训练，然后利用 `BP` 算法进行全局参数的微调，预估概率分布之间的交叉熵作为目标函数。在训练过程中，模型初始学习速率设置为 0.08，在前四个迭代训练过程中保持不变，后续训练时将学习率减半，当交叉验证值趋于收敛，停止训练。此外冲量值设置为 0.5，`Mini-batch` 大小设置为 256。训练完成后，利用 `BN` 特征训练得到的 `GMM-HMM` 模型进行识别解码。

### 3.5 BN+MFCC-GMM-HMM 模型

`BN+MFCC-GMM-HMM` 模型是使用提取的 `BN` 特征串接原始 `MFCC` 特征，通过 `LDA` 降维后训练得到 `GMM-HMM` 模型并进行识别解码。它训练所涉及的一些参数配置与 `BN-DNN` 训练一致。

### 3.6 实验结果及分析

根据上述实验在 `Kaldi` 上分别搭建基于 `MFCC`、`BN`、`BN+MFCC` 特征 `GMM-HMM` 声学模型语音识别系统以及 `DNN-HMM` 声学模型系统，对藏语进行语音识别性能测试。采用音节错误率作为系统性能的衡量指标。识别结果如表 2 所示。

表 2 模型识别结果

Table 2 Recognition results of models

模型类别	WER
MFCC-GMM-HMM	18.61%
DNN-HMM	15.77%
BN-GMM-HMM	15.98%
BN+MFCC-GMM-HMM	14.49%

以上结果显示瓶颈复合特征、DNN 后验特征以及单一瓶颈特征较基线模型 (MFCC-GMM-HMM) 在藏语连续语音识别音节错误率上分别下降了 4.12%、2.84%、2.63%，瓶颈复合特征具有最佳的识别表现，单一的瓶颈特征和深度神经网络后验特征识别表现相当。瓶颈特征不仅能够借助成熟的 GMM-HMM 技术训练声学建模，而且在融合了传统 MFCC 特征后，其又使系统识别率得到了更加明显的改善。

## 4 结语

本文研究了瓶颈特征及其复合特征在藏语拉萨话连续语音识别中的具体应用问题，并在 Kaldi 系统平台上进行相关实验，实验结果显示 BN 特征相比传统 MFCC 特征在藏语语音识别率上更有优势，而其复合特征相比 DNN 后验特征和单 BN 特征能够进一步降低藏语连续语音识别音节错误率，其不仅结合了 DNN 特征的语音长时相关性和高度抽象表征信号的能力，还结合了 MFCC 特征的具有人耳听觉的特性，利用成熟的 GMM-HMM 声学建模技术，在藏语拉萨话连续语音识别方面表现更好。下一步，我们将加大训练语料的时长，采用不同神经网络模型进行训练比对，以获得更高的系统识别率。

## 参考文献

- [1] Mohamed Abdel-rahman, Yu Dong, Li Deng. Investigation of full-sequence training of deep belief networks for speech recognition[C]//Eleventh Annual Conference of the International Speech Communication Association, 2010, 2846-2849.
- [2] Dahl George E, Yu Dong, Li Deng, et al. Large vocabulary continuous speech recognition with context-dependent DBN-HMMs[C]//Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on. IEEE, 2011: 4688-4691.
- [3] 王辉, 赵悦, 刘晓凤, 等. 基于深度特征学习的藏语语音识别[J]. 东北师大学报, 2015, 47(4):70-73.
- [4] 张宇聪. 基于深度学习的藏语拉萨方言语音识别的研究[D]. 兰州:西北师范大学物理与电子工程学院, 2016.
- [5] 袁胜龙, 郭武, 戴礼荣. 基于深度神经网络的藏语识别[J]. 模式识别与人工智能, 2015, 28(3):210-213.
- [6] Campbell William M, Douglas E. Sturim, Douglas A. Reynolds. Support vector machines using GMM supervectors for speaker verification[J]. IEEE signal processing letters, 2006, 13(5): 308-311.
- [7] Dardari Yu Dong, Michael L. Seltzer. Improved bottleneck features using pretrained deep neural networks[C]//Twelfth Annual Conference of the International Speech Communication Association. 2011., 237-240.
- [8] 麦提艾力.吐尔逊, 戴礼荣. 深度神经网络在维吾尔语大词汇量连续语音识别中的应用[J]. 数据采集, 2015, 30(2):365-371.
- [9] 刘学, 王松年, 郭武. 采用深层神经网络中间层特征的关键词识别[J]. 小型微型计算机系统, 2015, 36(7):1541-1544.
- [10] Hinton Geoffrey, Li Deng, Dong Yu, et al. Deep neural Networks for acoustic modeling in speech recognition: The shared views of four research groups[J]. IEEE Signal Processing Magazine, 2012, 29(6): 82-97.
- [11] Song Y, Jiang B, Bao Y B, et al. I-vector representation based on bottleneck features for language identification[J]. Electronics Letters, 2013, 49(24): 1569-1570.
- [12] Hinton Geoffrey. A practical guide to training restricted boltzmann machines[M]//Neural networks: Tricks of the trade. Springer Berlin Heidelberg, 2012, 599-619.
- [13] 吕丹桔, B. Hoffmeister. 汉语语音声学特征复合的研究[J]. 云南大学学报, 2010, 32(S1):368-371.
- [14] Gopinath Ramesh A. Maximum likelihood modeling with Gaussian distributions for classification[C]. Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on. IEEE, 1998, 2: 661-664.

- [15] 李冠宇, 孟猛. 藏语拉萨话大词表连续语音识别声学模型研究[J]. 计算机工程, 2012, 38(5):189-191.
- [16] 金鹏. 中国少数民族语言简志丛书藏语简志[M]. 北京: 民族出版社出版, 1983:1-198.
- [17] 拉龙东智. 藏语语音识别技术研究[D]. 拉萨: 西藏大学藏文信息技术研究中心, 2015.
- [18] MijitAblimit, Graham Neubig, Masato Mimura.Uyghur Morpheme-based Language Models and ASR[J]. ICSP. 2010:581-584.
- [19] Povey D, Burget L, Agarwal M, et al. The subspace Gaussian mixture model—A structured model for speech recognition[J]. Computer Speech & Language, 2011, 25(2): 404-439.
- [20] Yue Zhao, Rui Zhao, Xiaoyang Wang, Qiang Ji. Multilingual Articulatory Features Augmentation Learning[C]//23rd International Conference of Pattern Recognition, 2016, 2902-2906.
- [21] Grézl F, Karafiát M, Kontár S, et al. Probabilistic and bottleneck features for LVCSR of meetings[C]//Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on. IEEE, 2007, 4: IV-757-IV-760.