



## 基于瓶颈复合特征的声学模型建立方法

郑文秀, 赵峻毅, 文心怡, 姚引娣

(西安邮电大学 通信与信息工程学院, 西安 710121)

**摘 要:** 针对梅尔频率倒谱系数(MFCC)语音特征不能有效反映连续帧之间有效信息的问题, 基于神经网络相关性和紧凑性特征, 提出一种融合神经网络瓶颈特征与 MFCC 特征的复合特征构造方法, 提高语音的表征能力和建模能力。从语音数据中提取 MFCC 特征作为输入数据, 将 MFCC 特征和 BN 特征进行串接得到新的复合特征, 并进行 GMM-HMM 声学建模。在 TIMIT 数据库上的实验结果表明, 与单一的瓶颈特征和神经网络后验特征相比, 该方法识别率明显提升。

**关键词:** 神经网络; 梅尔频率倒谱系数; 瓶颈特征; 复合特征; 高斯混合模型-隐马尔科夫模型

开放科学(资源服务)标志码(OSID):



**中文引用格式:** 郑文秀, 赵峻毅, 文心怡, 等. 基于瓶颈复合特征的声学模型建立方法[J]. 计算机工程, 2020, 46(11):301-305, 314.

**英文引用格式:** ZHENG Wenxiu, ZHAO Junyi, WEN Xinyi, et al. Acoustic model construction method based on bottleneck compound feature[J]. Computer Engineering, 2020, 46(11):301-305, 314.

### Acoustic Model Construction Method Based on Bottleneck Compound Feature

ZHENG Wenxiu, ZHAO Junyi, WEN Xinyi, YAO Yindi

(School of Communication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China)

**[Abstract]** The Mel-Frequency Cepstral Coefficient (MFCC) speech features cannot effectively reflect the effective information between consecutive frames. To address the problem, this paper uses deep neural network to extract bottleneck features with long-term correlation and compactness of speech, and on this basis proposes a compound feature construction method that combines the neural network bottleneck features and the MFCC feature, so as to improve the speech characterization and modeling capabilities. The MFCC feature is extracted from the speech data as the input, and then concatenated with the BN feature to obtain a new compound feature. On this basis the acoustic modeling of Mixture Model-Hidden Markov Model (GMM-HMM) is implemented. Experimental results on the TIMIT database show that compared with the methods based on the single bottleneck feature and deep neural network posterior feature, the proposed method can significantly increase the recognition rate.

**[Key words]** Deep Neural Networks (DNN); Mel-Frequency Cepstral Coefficient (MFCC); bottleneck feature; compound feature; Gaussian Mixture Model-Hidden Markov Model (GMM-HMM)

**DOI:** 10.19678/j.issn.1000-3428.0056278

### 0 概述

近年来,随着语音识别技术的迅速发展,特别是深度神经网络(Deep Neural Networks, DNN)在大词汇量连续语音识别中的成功应用,使语音的识别正确率得到了很大的提升。语音识别系统一般包含特征提取、声学模型和解码识别3个部分<sup>[1]</sup>。特征提取是将原始数据中提取有利于后续过程中语音识别

的部分特征,消除大量冗余信息,对这些特征进行降维和去噪处理<sup>[2]</sup>。声学模型训练利用特征和标注训练模型来区分隐马尔科夫模型(Hidden Markov Model, HMM)状态,包括高斯混合模型(Gaussian Mixture Model, GMM)、深度信念网络(Deep Belief Networks, DBN)<sup>[3]</sup>、深度神经网络<sup>[4]</sup>和瓶颈-高斯混合模型(BottleNeck-GMM, BN-GMM)。

在 DNN-HMM 声学模型之前, GMM-HMM 声

**基金项目:** 国际科技合作项目一般项目“基于大数据信息决策的智慧农业自动灌溉系统研究”(2018KW-025)。

**作者简介:** 郑文秀(1979—),女,副教授、博士,主研方向为信息处理技术;赵峻毅、文心怡,硕士研究生;姚引娣,高级工程师、硕士。

**收稿日期:** 2019-10-14 **修回日期:** 2019-12-13 **E-mail:** 782933143@qq.com

学模型具有完善的理论知识体系,训练效率较高。传统的语音识别声学模型采用梅尔频率倒谱系数(Mel-Frequency Cepstral Coefficient, MFCC)特征对GMM-HMM进行建模。但是MFCC特征具有短时的特性,容易受到环境中噪声的影响,鲁棒性较差,忽略了连续帧之间的相关特性<sup>[5]</sup>。为利用GMM-HMM的性能优势,文献[6-7]研究了一种具有狭窄中间层的瓶颈深度神经网络,从神经网络的瓶颈层提取BN特征来代替MFCC特征,并应用于GMM-HMM声学模型训练,构造出BN-GMM-HMM,实验结果表明,该模型与DNN-HMM声学模型相比具有相当的识别能力。

本文将深度神经网络提取的语音长时相关性和紧凑性特征与传统MFCC特征相结合,构造一种新的复合特征参数流系统。该系统采用成熟的GMM-HMM声学模型,并利用复合特征参数流进行GMM-HMM模型的重构,以提升系统的识别率。

## 1 声学特征的提取

### 1.1 基于深度神经网络的瓶颈特征提取

2007年,GREZL等人<sup>[8]</sup>提出瓶颈的概念并在连续语音识别中成功应用。通过在BN-DBN引入瓶颈层减少了输出特征的维度,降低了后续的运算复杂度<sup>[9]</sup>。本文采用基于DNN模型的瓶颈声学特征,其训练与其他应用于语音识别的DNN训练程相似。输入数据为39维(12维滤波器输出值,1维对数能量和两者的一阶、二阶差分)、帧长为25ms、帧移为10ms、11帧拼接的MFCC特征<sup>[10]</sup>。

本文采用的DNN由输入层、5层隐层和输出层构成。相邻的两层神经元之间相互连接,层内神经元不连接<sup>[11]</sup>。基于BN-DNN的语音特征提取可以分为以下2个主要步骤:

**步骤1 建立DNN。**DNN的训练分成神经网络初始化和参数更新2个阶段,即预训练(Pre-training)和微调(Fine-tuning),建立一个训练好的DNN<sup>[12]</sup>。

1) Pre-training 阶段:采用贪婪算法进行非监督的方法训练一个受限玻尔兹曼机(Restricted Boltzmann Machine, RBM),再通过从底向上的方式训练多个RBM,底层RBM的输出值作为高层的输入值,获得一个DBN网络。

本文RBM的能量函数为高斯-伯努利型,可表示为:

$$E(v, h; \theta) = - \sum_{i=1}^I \sum_{j=1}^J \frac{w_{ij} h_j v_i}{\sigma_i} - \sum_{i=1}^I \frac{(v_i - b_i)^2}{2\sigma^2} - \sum_{j=1}^J a_j h_j \quad (1)$$

其中, $\theta$ 代表模型参数, $v_i$ 代表第*i*个显层神经元, $h_j$ 代表第*j*个隐层神经元, $w_{ij}$ 代表第*i*个神经元和第*j*个神经元的连接权重, $\sigma_i$ 代表控制能量宽度的参数, $b_i$ 代表显层神经元的第*i*个偏置, $a_j$ 代表隐层神经元的第*j*个偏置。

利用梯度下降(Gradient Descent, GD)算法对RBM的对数似然概率 $\log_a p(v, h; \theta)$ 进行计算,其推导公式为:

$$-\frac{\partial \log_a p(v, h; \theta)}{\partial \theta} = \left( \frac{\partial E_{\text{data}}(v, h)}{\partial \theta} \right) - \left( \frac{\partial E_{\text{model}}(v, h)}{\partial \theta} \right) \quad (2)$$

其中, $\left( \frac{\partial E_{\text{data}}(v, h)}{\partial \theta} \right)$ 表示对数据分布的期望, $\left( \frac{\partial E_{\text{model}}(v, h)}{\partial \theta} \right)$ 表示对模型的期望。由于显层数据的特征维度较高时,吉布斯采样(Gibbs Sampling)要采用很大的步数,因此采用与梯度相类似的通过*K*步(通常*K*=1)Gibbs的对比散度算法(Contrastive Divergence, CD)。由此,可得参数*W*、*b*、*a*的权值和偏置的更新公式为:

$$\Delta W_{ij} = \sum_{l=1}^m p(h_j = 1 | v^l) v^l - \tilde{v}^l p(h_j = 1 | v) \quad (3)$$

$$\Delta b_i = \sum_{l=1}^m p(v_i^l = 1 | h) - p(\tilde{v}_i^l = 1 | h) \quad (4)$$

$$\Delta a_j = \sum_{l=1}^m p(h_j = 1 | v^l) - p(h_j = 1 | \tilde{v}^l) \quad (5)$$

其中, $(\tilde{\phantom{x}})$ 表示由CD算法获得对*v<sub>i</sub>*的估计值,式(3)~式(5)是一个非监督的预训练过程,可将RBM调整到合适的初始值。

2) Fine-tuning 阶段:在最后一个RBM处采取反向传播(Back Propagation, BP)算法对整个神经网络自顶向下进行有监督地微调训练,完成DNN的建立。DNN的结构如图1所示。

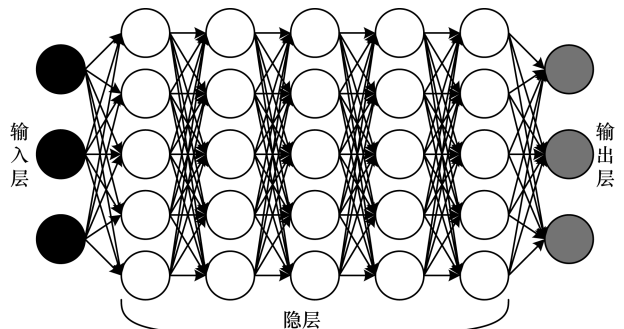


图1 DNN结构

Fig. 1 Structure of DNN

**步骤 2** 训练好 DNN 后, 将瓶颈层后面网络移除, 瓶颈层当作输出层获取 BN 特征<sup>[9]</sup>。BN-DNN 结构如图 2 所示。

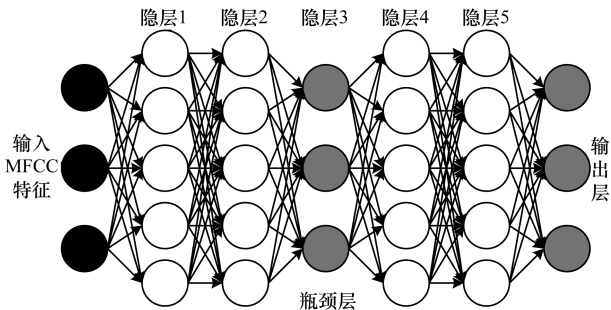


图 2 BN-DNN 结构

Fig. 2 Structure of BN-DNN

从图 2 可以看出<sup>[13]</sup>, 隐层 3 为瓶颈层, 将提取的 MFCC 声学特征作为输入数据经过显层和隐层对网络进行无监督预训练, 并采取 BP 算法对整个神经网络由后往前进行有监督地微调训练, 完成 DNN 的建立。训练好模型后, 将隐层 3 之后的隐层 4、隐层 5 及输出层去除, 并将瓶颈层作为输出层。DNN 具有强大的分类能力, 可以从数据中学习更有利于特定分类任务的特征表示, 因此提取出的语音瓶颈特征更有效。

## 1.2 新复合特征的构造

复合特征是指传统的短时特征和非短时差异特征复合后形成新的特征参数流。2010 年, 吕丹桔等人<sup>[14]</sup>提出将传统的短时特征如 MFCC/PLP 特征与采用 ANN 技术提取具有差异特征的 MLP 复合构成新的复合特征向量, 利用构造出的复合特征 GMM-HMM 声学建模在汉语的声学特征的识别率上比单特征有了明显提高。2018 年, 周楠等人<sup>[15]</sup>在藏语的研究中进行了瓶颈复合特征的相关实验, 结果表明, 复合特征相比于单 BN 特征和 DNN 后验特征系统, 识别率有了明显改善。本文采取 DNN 模型提取瓶颈特征与 MFCC 特征相结合构造出一个新的复合特征参数。

复合特征提取步骤如下:

- 1) 对语音的原始数据预处理得到 MFCC 特征。
- 2) 训练基于最大似然准则的 GMM-HMM 声学模型, 并通过区分性训练优化 GMM 模型。
- 3) 采用步骤 2 优化好的 GMM 模型对训练标注进行强制对齐, 获取帧级标注用于 DNN 训练。
- 4) 通过 Pre-training 初始化前面得到的 DNN 网络参数, 并根据交叉熵准则 Fine-tuning 训练 DNN 网络。

5) 移除上一步骤得到的 DNN 网络中瓶颈层之后的网络, 将瓶颈层作为输出层, 获取 BN 特征。

6) 将上述步骤获取的 BN 特征和 MFCC 特征进行串接获取复合特征。

复合特征的流程如图 3 所示。

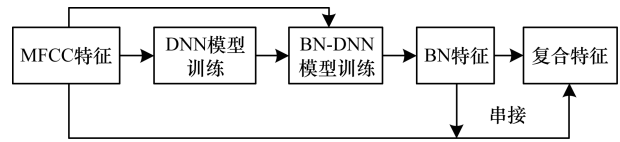


图 3 复合特征的训练流程

Fig. 3 Training procedure of compound features

## 2 仿真实验与结果分析

### 2.1 GMM-HMM 模型建立

本文采用三音素模型进行训练。三音素模型采用 A\_B\_C 形式, B 为当前状态, A 和 C 分别为前后状态, 对三音素单元使用自左向右的无状态间跨越的三状态 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<sup>[17]</sup>。

基于最大似然准则的 GMM-HMM 声学模型, 其输入为 39 维特征, 帧长为 25 ms, 帧移为 10 ms, HMM 中的每个状态设置 100 个独立的高斯分量。

### 2.2 瓶颈特征的 GMM-HMM 模型建立

由 DNN 作为特征提取模块提取出 BN 特征并使用 GMM-HMM 进行声学建模所构成的系统称为级联系统<sup>[18]</sup>。首先训练一个 DNN-bottleneck 神经网络提取瓶颈特征。输入层神经元数目 = 输入特征的帧数 × 每帧的维数, 即输入层节点  $429 = 39 \times 11$ ; 隐层中的神经元数目通常设置应尽量大, 使得它可以提升深度神经网络的性能, 且除 BN 层以外, 其他隐层的神经元数目相等; 瓶颈层的神经元数目通常设置和特征的单帧维数一样。因此, 本文的 BN-DNN 的结构表示为:  $429 - [1\ 024 - 1\ 024 - 39 - 1\ 024 - 1\ 024] - 429$ 。最后, 将得到的 BN 特征训练 GMM-HMM 模型, 进行识别解码。

### 2.3 复合特征的 GMM-HMM 模型建立

复合特征的 GMM-HMM 模型的训练采用提取的 39 维 BN 特征和 39 维传统 MFCC 特征进行串接, 得到 78 维高维度复合特征, 经过主成分分析法 (Principal Component Analysis, PCA) 降维<sup>[19-20]</sup>后, 获得 39 维的复合特征, 重复 2.2 节中复合特征提取步骤中的步骤 2 的训练方法, 训练复合特征的 GMM-HMM 声学模型并进行识别解码。复合特征的声学模型建立过程如图 4 所示。

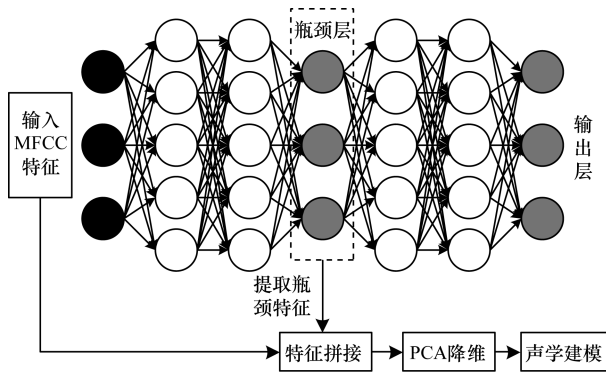


图4 复合特征的声学模型建立过程

Fig.4 Process of acoustic model establishing with compound features

## 2.4 实验过程

实验过程如下:

### 1) 数据集

本文采用 TIMIT 声学-音素学连续语音语料库来验证实验的有效性,共包含 4.3 h 的语音数据(其中包含 1.1 h 的 NIST 测试数据)。数据集中训练集包含 462 个不同发音人朗读的 4 620 个句子,测试集包含 162 个不同发音人朗读的 1 620 个句子,且训练集与测试集没有同一个说话人。

### 2) 参数的设置

在 Pre-training 阶段,所有的 RBM 使用 CD 算法配合小批量随机梯度下降法进行训练,每一个 mini-batch 的大小为 128,冲量因子设置为 0.9,不使用权重衰减,学习率训练 200 代 RBM。在 Fine-tuning 阶段,所有的参数设置和预训练阶段相同。

### 3) 评估标准

本文将词错误率(Word Error Rate, WER)作为评估标准。WER 的公式为:

$$WER = \frac{S + D + I}{N} \times 100\% \quad (6)$$

其中,  $S$  为替换数,  $D$  为删除数,  $I$  为插入数,  $N$  为单词数。

## 2.5 结果分析

### 实验 1 隐层中神经元个数的最优设置

在实验 1 中,本文通过调整 BN-DNN 模型中隐层的神经元个数确定最佳的网络结构配置。隐层依次设置为 256、512、1 024、2 048 这 4 种情况,每个隐层神经元个数相同,且 BN 层神经元数目临时设置为 39。不同神经元个数的模型性能比较如表 1 所示。

表 1 不同神经元个数的复合特征声学模型性能比较

Table 1 Performance comparison of compound features acoustic models with different numbers of neurons %

隐层神经元个数	词错误率
256_BN + MFCC-GMM-HMM_39	15.46
512_BN + MFCC-GMM-HMM_39	14.38
1 024_BN + MFCC-GMM-HMM_39	13.75
2 048_BN + MFCC-GMM-HMM_39	13.87

从表 1 可以看出,在 DNN 为五层隐层的情况下,除瓶颈层神经元数目为 39 外,其他隐层均为 1 024 时,新的复合特征建立的声学模型效果达到最佳。从而验证在 2.2 节提出的深度神经网络模型中,隐层神经元数目尽量大可以提升深度神经网络,在神经元数目达到 1 024 时,WER 达到最小值 13.75%,神经元数目继续增加,WER 反而增加。所以,神经元数目并不是越大越好,五层隐层效果最好的神经元数目为 1 024。

### 实验 2 BN 层中神经元个数的最优设置

根据实验 1 确定隐层神经元数目为 1 024,改变 BN 层中神经元的个数,分别设置为 20、30、39 及 40 这 4 种情况来确定 BN 层为 39 个神经元时是否为最优参数。不同 BN 层的模型性能比较如表 2 所示。

表 2 不同 BN 层的复合声学模型性能比较

Table 2 Performance comparison of compound acoustic models of different BN layers %

BN 层神经元个数	词错误率
1 024_BN + MFCC-GMM-HMM_20	13.86
1 024_BN + MFCC-GMM-HMM_30	13.81
1 024_BN + MFCC-GMM-HMM_39	13.75
1 024_BN + MFCC-GMM-HMM_50	13.82

从表 2 可以看出,在 BN 层数改变的 4 种情况下,WER 并没有太大的改变,在最大差距数目为 39 时,比 20 降低了 0.11%,在最小差距数目为 39 时,比 30 降低了 0.06%,但验证了 BN 层神经元数目为 39 时,根据新的复合特征建立的 GMM-HMM 声学模型性能最好。

### 实验 3 4 种不同特征的声学模型识别率的对比

MFCC 特征、BN 特征及新的复合特征进行 GMM-HMM 声学建模以及深度神经网络后验特征的识别结果对比如表 3 所示。

表 3 不同模型识别的词错误率

Table 3 Word error rate recognized by different models %

模型类别	词错误率
MFCC-GMM-HMM	19.38
1 024_DNN-HMM	17.31
1 024_BN-GMM-HMM_39	17.42
1 024_BN + MFCC-GMM-HMM_39	13.75

从表 3 可以看出,与其他语音特征相比,基于深度神经网络的瓶颈特征与传统 MFCC 特征的复合特征具有最佳的识别效果。相比单一的 MFCC 特征 WER 下降 5.63%,与单一的瓶颈特征相比,深度神经网络后验特征具有相当的识别性能。而复合特征

的识别效果相比于神经网络后验特征和单一瓶颈特征分别提高了 3.56% 和 3.67%。这是因为当训练数据与测试数据相匹配时,使用 BN 特征相比于 MFCC 特征能获得更低的错误率,但是当训练数据与测试数据不匹配时,BN 特征与 MFCC 特征复合才能获得更好的识别率。

### 3 结束语

本文针对 TIMIT 语音数据集连续语音识别,将语音数据中提取到的 MFCC 特征作为输入数据,经过最大似然准则训练 GMM-HMM 声学模型,并根据区分性训练优化 GMM 模型。通过预训练初始化参数并根据交叉熵准则微调训练 DNN 网络,移除训练好的 DNN 网络瓶颈层之后的网络,从而获取具有更强区分性的 BN 特征。将 MFCC 特征和 BN 特征进行串接,建立复合特征的 GMM-HMM 声学模型。实验结果表明,神经网络提取的 BN 特征在识别效果上比传统的 MFCC 特征更具优势,两者串接而成的复合特征的声学模型在识别率上有了大幅提升,从而验证了本文声学建模方法较优的识别性能。下一步拟将不同的神经网络模型与传统的 MFCC 特征进行对比,构建复合的声学模型,以期获得更好的识别率。

### 参考文献

- [ 1 ] SCHWARZ P. Phoneme recognition based on long temporal context [ EB/OL ]. [ 2013-09-10 ]. [http://speech. fit. Vutbr. cz/soft-ware/Phoneme-recognizer-based-long-temporal-context](http://speech.fit.vutbr.cz/soft-ware/Phoneme-recognizer-based-long-temporal-context).
- [ 2 ] WANG Yi, YANG Junan, LIU Hui, et al. Bottleneck feature extraction method based on hierarchical deep sparse belief network [ J ]. Pattern Recognition and Artificial Intelligence, 2015, 28(2) : 173-180. ( in Chinese )  
王一,杨俊安,刘辉,等.基于层次稀疏 DBN 的瓶颈特征提取方法 [ J ]. 模式识别与人工智能, 2015, 28(2) : 173-180.
- [ 3 ] MOHAMED A R, DAHL G, HINTON G. Acoustic modeling using deep belief networks [ J ]. IEEE Transactions on Audio, Speech and Language Processing, 2012, 20(1) : 14-22.
- [ 4 ] DAHL G E, YU D, DENG L, et al. Context-dependent pre-trained deep neural networks for large vocabulary speech recognition [ J ]. IEEE Transactions on Audio, Speech and Language Processing, 2012, 20(1) : 30-42.
- [ 5 ] LUO Yuan, LIU Yu, ZHANG Yi, et al. Speech bottleneck feature extraction method based on overlapping group lasso sparse deep neural network [ J ]. Journal of Speech Communication, 2018, 99 : 56-61.
- [ 6 ] YU D, SELTZER M. Improved bottleneck features using pre-trained deep neural networks [ C ] // Proceedings of the 12th Annual Conference of the International Speech Communication Association. Florence, Italy : [ s. n. ], 2011 : 237-240.
- [ 7 ] LIU Yuqing, WANG Tianhao, XU Xu. New adaptive activation function for deep learning neural networks [ J ]. Journal of Jilin University Science Edition, 2019, 57(4) : 857-859. ( in Chinese )  
刘宇晴,王天昊,徐旭.深度学习神经网络的新型自适应激活函数 [ J ]. 吉林大学学报(理学版), 2019, 57(4) : 857-859.
- [ 8 ] GREZL F, KARAFIATT M, KONTAR S, et al. Probabilistic and bottle-neck features for LVCSR of meetings [ C ] // Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. Honolulu, USA : IEEE Press, 2007 : 757-760.
- [ 9 ] CHEN Lei, YANG Junan, WANG Yi, et al. A feature extraction method based on discriminative and adaptive bottleneck deep belief network in large vocabulary continuous speech recognition system [ J ]. Journal of Signal Processing, 2015, 31(3) : 290-298. ( in Chinese )  
陈雷,杨俊安,王一,等. LVCSR 系统中一种基于区分性和自适应瓶颈深度置信网络的特征提取方法 [ J ]. 信号处理, 2015, 31(3) : 290-298.
- [ 10 ] SINISCALCHI S M, YU D, DENG L, et al. Speech recognition using long-span temporal patterns in a deep network Model [ J ]. IEEE Signal Processing Letters, 2013, 20(3) : 201-204.
- [ 11 ] WANG Zhaokai, LI Yaxing, FENG Xupeng, et al. Personalized information recommendation based on deep belief network [ J ]. Computer Engineering, 2016, 42(10) : 201-206. ( in Chinese )  
王兆凯,李亚星,冯旭鹏,等.基于深度信念网络的个性化信息推荐 [ J ]. 计算机工程, 2016, 42(10) : 201-206.
- [ 12 ] LI Jinhui, YANG Junan, WANG Yi. New feature extraction method based on bottleneck deep belief networks and its application in language recognition [ J ]. Computer Science, 2014, 41(3) : 263-266. ( in Chinese )  
李晋徽,杨俊安,王一.一种新的基于瓶颈深度信念网络的特征提取方法及其在语种识别中的应用 [ J ]. 计算机科学, 2014, 41(3) : 263-266.
- [ 13 ] PINTO J, SIVARAM G S V S, DOSS M M, et al. Analysis of MLP based hierarchical phoneme posterior probability estimator [ J ]. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 19(2) : 225-241.
- [ 14 ] LÜ D, HOFFMEISTER B. Study on the compounding of Chinese phonetic acoustics [ J ]. Journal of Yunnan University ( Natural Science Edition ), 2010, 32( Sup ) : 368-371. ( in Chinese )  
吕丹桔, HOFFMEISTER B. 汉语语音声学特征复合的研究 [ J ]. 云南大学学报(自然科学版), 2010, 32(增刊) : 368-371.

(上接第 305 页)

- [15] ZHOU Nan, ZHAO Yue, LI Yaoqiang, et al. Study on continuous speech recognition based on bottleneck features for Lhasa-Tibetan dialect[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2018, 54(2): 249-254. (in Chinese)  
周楠, 赵悦, 李要强, 等. 基于瓶颈特征的藏语拉萨话连续语音识别研究[J]. 北京大学学报(自然科学版), 2018, 54(2): 249-254.
- [16] LIU Diyuan. Research on BN features based acoustic modeling and its application in keyword retrieval[D]. Hefei: University of Science and Technology of China, 2015. (in Chinese)  
刘迪源. 基于 BN 特征的声学建模研究及其在关键词检索中的应用[D]. 合肥: 中国科学技术大学, 2015.
- [17] LI Yunhong, LIANG Sicheng, JIA Kaili, et al. An improved speech recognition base on DNN-HMM model[J]. Journal of Applied Acoustics, 2019, 38(3): 371-377. (in Chinese)  
李云红, 梁思程, 贾凯莉, 等. 一种改进的 DNN-HMM 的语音识别方法[J]. 应用声学, 2019, 38(3): 371-377.
- [18] QIN Chuxiong, ZHANG Lianhai. Deep neural network base feature extraction for low-resource speech recognition[J]. Acta Automatica Sinica, 2017, 43(7): 1208-1219. (in Chinese)  
秦楚雄, 张连海. 基于 DNN 的低资源语音识别特征提取技术[J]. 自动化学报, 2017, 43(7): 1208-1219.
- [19] LI Haixia, WU Suyi. Attribute reduction and optimization for massive seismic data base on principal component analysis[J]. China Earthquake Engineering Journal, 2019, 41(3): 757-762. (in Chinese)  
李海霞, 吴苏怡. 基于主成分分析法的海量地震数据属性降维优化[J]. 地震工程学报, 2019, 41(3): 757-762.
- [20] ZHOU Bufang. A dissertation for the master degree of engineering[D]. Zhangzhou: Minnan Normal University, 2017. (in Chinese)  
周步芳. 多标签学习的特征降维方法[D]. 漳州: 闽南师范大学, 2017.

编辑 索书志