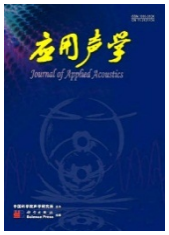


应用声学
Journal of Applied Acoustics
ISSN 1000-310X,CN 11-2121/O4

《应用声学》网络首发论文

题目：一种改进的 DNN-HMM 的语音识别方法
作者：李云红，梁思程，贾凯莉，张秋铭，宋鹏，何琛，王刚毅，李禹萱
收稿日期：2018-09-15
网络首发日期：2019-03-01
引用格式：李云红，梁思程，贾凯莉，张秋铭，宋鹏，何琛，王刚毅，李禹萱．一种改进的 DNN-HMM 的语音识别方法[J/OL]．应用声学.
http://kns.cnki.net/kcms/detail/11.2121.O4.20190228.1539.002.html

一种改进的 DNN-HMM 的语音识别方法*

李云红[1†] 梁思程[1] 贾凯莉[1] 张秋铭[1]
宋 鹏[1] 何 琛[1] 王刚毅[1] 李禹萱[2]

(1 西安工程大学电子信息学院 西安 710048)
(2 国家电网西安供电公司 西安 710032)

摘要 针对深度神经网络与隐马尔可夫模型(DNN-HMM)结合的声学模型在语音识别过程中建模能力有限等问题，提出了一种改进的 DNN-HMM 模型语音识别算法。首先根据深度置信网络(DBN)结合深度玻尔兹曼机(DBM)，建立深度神经网络声学模型，然后提取梅尔频率倒谱系数(MFCC)和对数域的 Mel 滤波器组系数(Fbank)作为声学特征参数，通过 TIMIT 语音数据集进行实验。实验结果表明：结合了 DBM 的 DNN-HMM 模型相比 DNN-HMM 模型更具优势，其中，使用 MFCC 声学特征在词错误率与句错误率方面分别下降了 1.26%和 0.20%。此外，使用默认滤波器组的 Fbank 特征在词错误率与句错误率方面分别下降了 0.48%和 0.82%，并且适量增加滤波器组可以降低错误率。总之，研究取得句错误率与词错误率分别降低到 21.06%和 3.12%的好成绩。

关键词 语音识别，深度神经网络，声学模型，声学特征
中图法分类号：TN912.34 **文献标志码**：A

An improved speech recognition method based on DNN-HMM model

LI Yunhong[1] LIANG Sicheng[1] JIA Kaili[1] ZHANG Qiuming[1] SONG Peng[1]
HE Chen[1] WANG Gangyi[1] LI Yuxuan[2]

(1 School of Electronics and Information, Xi'an Polytechnic University, Xi'an 710048, China)
(2 State Grid Xi'an Power Supply Company, Xi'an 710032, China)

Abstract The acoustic model combined with deep neural network and hidden Markov model (DNN-HMM) has been used extensively in today's speech recognition system. In this paper, an improved DNN-HMM model speech recognition algorithm is proposed. First, a deep neural network acoustic model is built by the deep belief network (DBN) and the deep Boltzmann machine (DBM). Then the Mel frequency cepstral coefficient (MFCC) and the log filter coefficient of the log domain (Fbank) are extracted as an acoustic feature parameter. Finally, the experiment is performed on the TIMIT speech data set. The experimental results show that the DNN-HMM model combined with DBM has more advantages than DNN-HMM model, in which the MFCC acoustic features can reduce the word error rate and sentence error rate by 1.26% and 0.20% respectively. Moreover, using the Fbank feature default filter group rate decreases the word error rate and sentence error rate by 0.48% and 0.82% respectively, and an appropriate increase in the filter bank group can reduce the error rate. In brief, the sentence error rate and the word error rate are reduced to 21.06% and 3.12% respectively.

Key words Speech recognition, Deep neural network, Acoustic model, Acoustic feature

2018-09-15 收稿; 2018-12-31 定稿
*国家自然科学基金资助项目(61471161)，陕西省科技厅自然科学基础研究重点项目(2016JZ026)，国家级大学生创新创业项目(201810709009)
作者简介：李云红(1974-)，女，辽宁锦州人，博士，教授，研究方向：信号与信息处理。
†通讯作者 E-mail: hitliyunhong@163.com

0 引言

声学模型作为语音识别系统的主要模型之一，利用一系列声学特征完成建模训练，能够明确各声学基

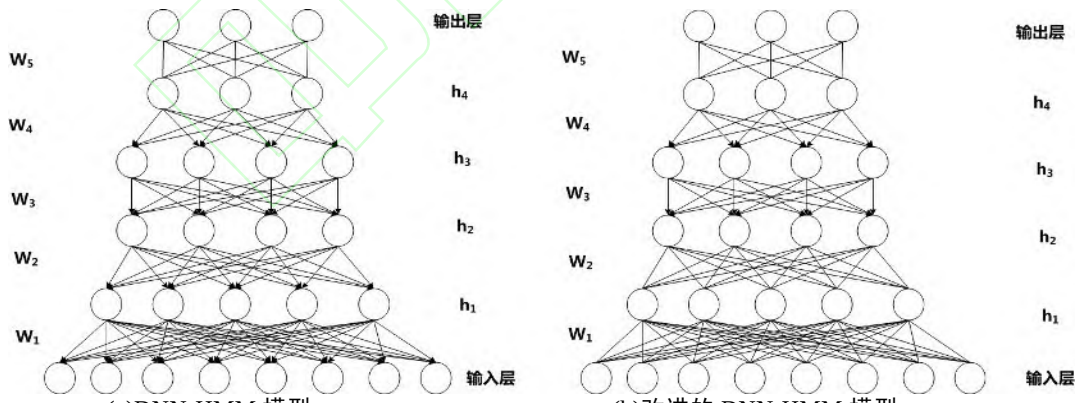
元相关发音模式。目前广泛应用的声学建模研究主要围绕高斯混合模型-隐马尔可夫模型(Gaussian mixture model- hidden Markov model, GMM-HMM)[1]展开。胡政权等[2]提出了梅尔频率倒谱系数(Mel-frequency cepstral coefficients, MFCC)参数提取的改进方法。赵涛涛等[3]提出了经验模态分解和加权 Mel 倒谱的语音共振峰提取算法。但是，随着深度学习在词识别率方面取得跨越性突破后，应用它建立声学模型成为了研究人员关注的焦点[4-8]。

2000 年，深度学习领域的专家 Hinton 等[9]提出了限制玻尔兹曼机(Restricted Boltzmann machine, RBM)，这种模型结构是可见层节点与隐藏层节点全部连接，相同层节点之间互相独立。2006 年，Hinton 等提出了基于层叠的 RBM 算法，即深度置信网络(Deep belief networks, DBN)，表明了深层神经网络模型在特征提取以及模型表达方面具有优异的表现。Mohamed 等[10]首次使用 DBN 来取代传统的 GMM 来为 HMM 状态输出特征分布建模，并成功搭建 DBN-HMM 声学模型应用于一个单音素识别系统，通过实验表明在词错误率方面下降到了 20.3%。最近几年，国内外专家学者在声学特征方面进行深入研究，使得深度学习理论在语音识别领域再次有了进一步的发展。张劲松等[11]比较了几种不同特征对识别率的影响，使用 Mel 滤波器组系数(Mel-scale filter bank, Fbank)作为声学特征，具有更好的识别率。Kovacs 等[12]更是在 Fbank 特征基础上利用自回归的方法来调整模型的鲁棒性，取得了较好的识别结果。

理论方面，经过多年研究发展，深度学习理论与语音识别技术的结合[13-14]已然达到较为成熟的阶段；应用方面，从最初的人工神经网络(Artificial neural network, ANN)到现在的深层神经网络(Deep neural network, DNN)，可以说神经网络已经达到实际应用阶段[15]。Salakhutdinov 等[16]提出的深度玻尔兹曼机(Deep Boltzmann machine, DBM)以 RBM 为基础，模型中单元各层均为无向连接，使模型处理不确定样本的健壮性更强。基于此，论文结合 DBM，在 Kaldi 平台上建立改进的 DNN-HMM 语音识别模型[17]，经语音识别库 TIMIT 的测试实验，取得了较好的语音识别结果。

1 改进的 DNN-HMM 声学模型

DNN-HMM 声学模型是由 DBN 模型组成的深度神经网络，DBN 模型隐藏层采用 RBM 组成的有向图模型。而改进的 DNN-HMM 声学模型由 DBM 模型和 DBN 模型混合而成。模型结构对比如图 1 所示。DBM 模型是由两层 RBM 组成的无向图模型，每层节点的采样值均由两层连接的节点共同计算。但是 DBM 模型训练时间长度与它的层数和每层的节点数有关。DBN 模型是由四层 RBM 组成的有向图模型，在预训练过程中，上层是输出，下层是输入。所有层训练完毕后，由最上层开始向下进行有监督微调。



(a)DNN-HMM 模型　　(b)改进的 DNN-HMM 模型

图 1 模型结构

Fig.1 Model structure

如图 1 所示，DNN-HMM 模型和改进的 DNN-HMM 模型都有 1 个输入层，4 个隐藏层，1 个输出层。h_1、h_2、h_3、h_4 分别对应 4 个隐藏层，W_1、W_2、W_3、W_4、W_5 分别对应层间的连接权重。模型相同层节点不连接，不同层节点之间全部连接。DNN-HMM 模型输入层、h_1、h_2、h_3、h_4 之间是有向图全连接的 DBN 模型。改进的 DNN-HMM 模型的输入层、h_1、h_2 之间是无向图全连接的 DBM 模型，h_2、h_3、h_4 之间是有向图全连接的 DBN 模型。固定长度的向量作为模型输入，改进的 DNN-HMM 模型先由 h_1、h_2 训练，h_2 作为 DBM 模型的输出层，同时也是 h_3、h_4 的输入，输出是当前输入信息的特征表示。

RBM 是基于能量的模型，可以捕获变量的相关性。其定义为

$$E(v,h)=-\sum_{}^{n}\sum_{}^{m}w_{ij}h_iv_j-\sum_{}^{m}b_jv_j-\sum_{}^{n}c_ih_i\,, \tag{1}$$

公式(1)表示每一个可视节点与隐藏节点之间构成的能量函数。其中，m 是可视节点的个数，n 是隐藏节点的个数，b 、c 是可视层和隐藏层的偏置。由于 RBM 目标函数要累加所有可视层和隐藏层节点取值的能量，其计算也面临指数级的复杂度。因此，将计算能量累加转换为求解概率的问题，即得到的 v,h 的联合概率为

$$p(v,h)=\frac{e^{-E(v,h)}}{\Sigma_{v,h}e^{-E(v,h)}}\,. \tag{2}$$

通过公式(2)简化能量函数的求解，使得求解的能量值最小。由统计学的一个理论，能量低发生的概率大，因此引入自由能量函数最大化联合概率，公式如下：

$$FreeEnergy(v)=-In\Sigma e^{-E(v,h)}\,, \tag{3}$$

$$P(v)=\frac{e^{-FreeEnergy(v)}}{Z}\,,Z=\Sigma_{v,h}e^{-E(v,h)}\,, \tag{4}$$

其中，Z 是归一化因子，故联合概率可以表示为

$$Inp(v)=-FreeEnergy(v)-InZ\,, \tag{5}$$

公式(5)中等号左边是似然函数 $p(v)$，右边第一项是整个网络自由能量总和的负值。

整个深度神经网络模型应用误差反向传播算法，让目标函数获得最优值，从而达到训练目的。针对深度神经网络进行训练时，目标函数通常替换为交叉熵，在实际优化阶段，使用随机梯度下降法来处理。换言之，对于多状态分类问题中目标函数往往使用取负值的对数概率，如公式(6)所示：

$$F_{CE}=-\sum_{u=1}^{U}\sum_{t=1}^{T_u}\log y_{ut}(s_{ut})\,, \tag{6}$$

其中，s_{ut} 是 t 时刻的状态，F_{CE} 为状态标签与预测状态分布 $y(s)$ 之间的交叉熵。目标函数与输入 $a_{ut}(s)$ 间的梯度可以记为

$$\frac{\partial F_{CE}}{\partial a_{ut}(s)}=-\frac{\partial\log y_{ut}(s_{ut})}{\partial a_{ut}(s)}=y_{ut}(s)-\delta_{ss_{ut}}\,, \tag{7}$$

公式(7)中 $\delta_{ss_{ut}}$ 是克罗内克函数，满足：

$$\delta_{ss_{ut}}=\begin{cases}1 & s=s_{ut}\\ 0 & s\neq s_{ut}\end{cases}\,, \tag{8}$$

由公式(8)，网络参数的调整方法使用反向传播算法。

改进的 DNN-HMM 模型与 DNN-HMM 模型不同的是底层使用了 DBM 模型对输入的语音信号进行了处理。DBM 模型中每一个隐藏节点的状态都由它直接连接的上下层节点共同计算决定，因此相比 DNN-HMM 模型可以对输入的语音信号进行更好的降维，捕捉不同语音的特征。同时，高层采用 DBN 模型结构避免了 DBN 模型开始训练时容易过拟合的现象，保持了良好的性能。

2 Fbank 特征

在语音识别领域当中，使用对角协方差矩阵的 GMM，将 MFCC 作为声学特征一直是研究的常用手法。MFCC 声学特征的计算过程如图 2 所示。

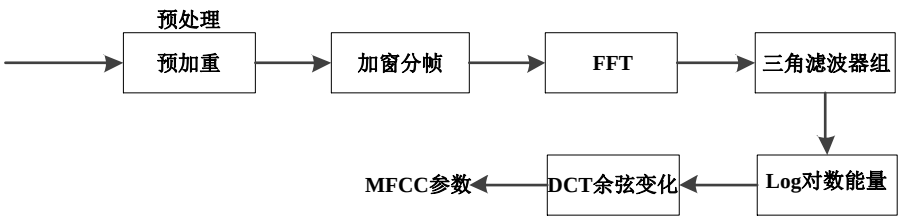


图 2 MFCC 计算流程图
Fig.2 MFCC flow chart

如图 2 所示，经预处理和快速傅里叶变换(Fast Fourier transformation, FFT)得到语音信号各帧数据的频谱参数，通过一组 N 个三角带滤波器构成的 Mel 频率滤波器作卷积运算，然后对输出的结果作对数运算，依次得到对数能量 $S(m), m=1,2,3,\cdots,N$ ，最后经离散余弦变换(Discrete cosine transform, DCT)，得到 MFCC 参数，如公式(9)所示：

$$C_i(n)=\sum_{m=1}^{M}S(m)\cos[\frac{\pi n(m-0.5)}{M}],0\le m\le M\ , \tag{9}$$

其中， n 代表 MFCC 声学特征的个数， $C_i(n)$ 是第 i 帧的第 n 个 MFCC 系数，作为 Log 对数能量模块的输出， M 是 Mel 滤波器的个数。

Fbank 声学特征省略了 MFCC 声学特征提取过程的 DCT 模块，将 Log 对数能量模块的输出直接作为输入语音的声学特征。在三角滤波器组模块，使用 N 个三角带滤波器就可以得到 N 维相关性较高的 Fbank 特征。而经过 DCT 计算提取的 MFCC 特征，将能量集中在低频部分，具有更好的判别度。

因此，使用 GMM 进行语音识别时，由于 GMM 忽略不同特征维度的相关性，MFCC 特征更加适合。而基于深度神经网络的语音识别中，深度神经网络可以更好地利用 Fbank 特征相关性较高的特点，降低语音识别的词错误率。另外，Fbank 声学特征相比 MFCC 声学特征，减小了声学特征提取时的计算量，容易进行带宽调节，得到最佳带宽的识别结果，从而进一步提高语音识别的正确率。

3 实验过程与结果分析

3.1 实验过程

3.1.1 GMM-HMM 声学模型的建立

(1)特征提取

实现帧长 25 ms、帧移 10 ms、特征维度 39 维(12 维输出、1 维对数能量及两者一阶、二阶差分)的 MFCC 特征的提取，然后进行倒谱均值方差归一化的处理。

(2)训练 GMM-HMM 模型

在模型训练过程中考虑将上下文相关的三音素融入声学模型，并以此作为声学基元进行模型训练，最后将训练后的模型输出特征进行解码。

在 Kaldi 开发平台中，三音素模型采用 A_B_C 结构形式，其中 B 为当前状态，A 和 C 为上下文。训练过程如表 1 所示。首先进行单音素模型训练，并按照设置的次数对数据对齐，然后以单音素模型为输入训练上下文相关的三音素模型并实现数据对齐，接下来对特征使用线性区分分析(Linear discriminant analysis, LDA)和最大似然线性回归(Maximum likelihood linear transform, MLLT)进行变换并训练加入 LDA 和 MLLT 的三音素模型，最后进行说话人自适应训练(Speaker adaptive training, SAT)得到 LDA+MLLT+SAT 的三音素模型，整个过程逐步实现了特征参数的优化。

表 1　基础模型训练过程

Table 1　Basic model training process

模型	解释
Mono	单音素模型
Mono_ali	单音素对齐
Tri1	三音素模型
Tri1_ali	三音素对齐
Tri2b	LDA+MLLT 特征的三音素模型
Tri2b_ali	LDA+MLLT 特征的三音素对齐
Tri3b	LDA+MLLT+SAT 特征的三音素模型
Tri3b_ali	LDA+MLLT+SAT 特征的三音素对齐

最后对识别结果进行强制性对齐，获得聚类后每个三音素的状态号来作为深度神经网络训练调谐时候

的标签信息，并以此作为训练 DNN 模型和改进的 DNN 模型的基础模型。

3.1.2 深度神经网络声学模型的建立

(1)监督信息的生成

因为 RBM 模型训练不适用不同长度的语音音素，论文通过强制对齐 GMM-HMM 基线系统识别结果，得到各聚类三音素状态，即模型 DNN 和改进模型 DNN 网络调参过程中所需标签信息。

(2)特征提取过程

在进行深度神经网络模型训练时，使用基于 MFCC 与 Fbank 两种不同的声学特征完成训练与解码，同时变更 Fbank 特征下滤波器组数量，观察不同滤波器组数量的 Fbank 特征对 DNN 和改进模型 DNN 网络识别结果的影响。

(3)网络参数设定

整个深度神经网络模型包含 1 个输入层、4 个隐藏层和 1 个输出层，网络输入选择超长帧(连续 11 帧组成)，隐藏层共有 1024 个节点，输出层共有 1366 个节点，各节点关联各种音素标签，输出层用 Softmax 网络作分类。

另外，由于深度神经网络模型参数调谐过程中需要根据开发集和测试集识别率的对比控制迭代次数。故在训练集中选取 3000 条语句作为开发集，选择 1000 条语句构成测试集。

(4)网络训练

首先初始化参数，设置 RBM 模型迭代 20 次。设置最小交叉熵为目标函数，借此调整参数。通过开发集与测试集测试得到识别准确率与迭代次数关系如图 3 所示。

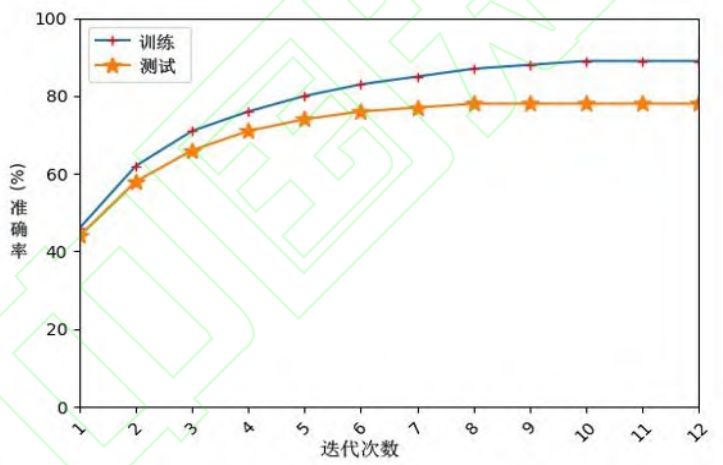


图 3 预测准确率与迭代次数变化

Fig.3 Prediction accuracy and number of iterations

(5)深度神经网络声学模型

结合深度神经网络输出层节点的输出值计算状态输出的后验证概率，调用 Kaldi 中的 nnet-forward 工具进行解码识别。

3.2 实验结果

根据上述步骤在 Kaldi 语音识别系统开发平台上训练单音素模型，并在此模型上优化训练三音素模型作为深度神经网络训练的基础模型。以训练好的三音素基础模型对分别使用 MFCC 特征和 Fbank 特征的模型进行训练解码。

整个实验中分别使用了滤波器组数目为 8、19、30、41、52、70、81 的 Fbank 特征对 DNN-HMM 模型和改进的 DNN-HMM 进行建模，Fbank 特征滤波器组数初始值设为 8，实验中首先对 8 组滤波器的 Fbank 特征进行训练解码，然后修改滤波器组数目进一步实验分析，比较滤波器组数目对实验结果的影响。

一个音素的发音时间一般在 9 帧左右，拼接特征的选择在 9 帧以上。实验中，拼接特征选择 11 帧，左右各 5 帧。根据 Fbank 特征滤波器组数目的不同，输入层节点个数分别设置为 88、209、330、451、572、770、891。经训练误差的比较后，4 个隐藏层节点个数选择 1024。输出层 1366 个节点，关联各种音素标签。

MFCC 特征下 GMM-HMM、DNN-HMM 和改进的 DNN-HMM 声学模型的句错误率与词错误率如表 2 所示。改进的 DNN-HMM 声学模型在不同 Fbank 特征下的句错误率和词错误率如表 3 所示，与 DNN-HMM 的识别率比较如图 4 所示。

表 2　MFCC 特征下声学模型的识别率

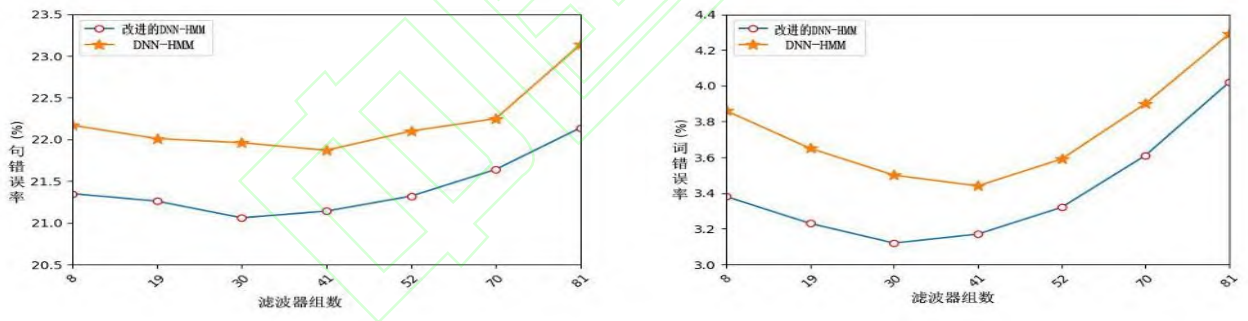
Table 2　Recognition rate of acoustic model under MFCC characteristics

声学模型	句错误率 SER	词错误率 WER
GMM-HMM	28.48%	6.13%
DNN-HMM(MFCC)	23.63%	4.35%
改进的 DNN-HMM(MFCC)	22.37%	4.15%

表 3 改进的 DNN-HMM 在不同 Fbank 特征下的识别率

Table 3　Recognition rate of improved DNN-HMM under different Fbank features

声学模型	句错误率 SER	词错误率 WER
改进的 DNN-HMM(8Filter-bank)	21.35%	3.38%
改进的 DNN-HMM(19Filter-bank)	21.26%	3.23%
改进的 DNN-HMM(30Filter-bank)	21.06%	3.12%
改进的 DNN-HMM(41Filter-bank)	21.14%	3.17%
改进的 DNN-HMM(52Filter-bank)	21.32%	3.32%
改进的 DNN-HMM(70Filter-bank)	21.64%	3.61%
改进的 DNN-HMM(81Filter-bank)	22.14%	4.02%



(a)句错误率影响　　(b)词错误率影响

图 4 改进的 DNN-HMM 与 DNN-HMM 模型错误率比较

Fig.4 Comparison of error rates between improved DNN-HMM and DNN-HMM models

3.3 实验分析

(1)根据表 2 的结果可以确定，在 MFCC 声学特征下，与传统 GMM-HMM 方法、DNN-HMM 方法相比较，改进的 DNN-HMM 声学建模方法在句错误率与词错误率方面均有下降，分别为 22.37%和 4.15%。这表明后者在声学建模方面相比 DNN 模型、GMM 模型对于复杂的语音数据有着更强的建模能力。

(2)从表 3 可以看出，滤波器组数量不断增多时，改进的 DNN-HMM 模型得到的句错误率与词错误率呈现先降后增的趋势。说明适当的增加滤波器组数量可以使识别结果更好，但是当增加到一定数量时结果反而会下降。论文实验中，滤波器组数量为 30 时，句错误率与词错误率达到最小值，分别为 21.06%和 3.12%。

(3)从图 4 可以看出，改进的 DNN-HMM 声学模型比 DNN-HMM 声学模型在不同滤波器组数量时句错误率与词错误率均有所下降，其中在滤波器组数量为默认值时，句错误率下降了 0.48%，词错误率下降了 0.82%。说明了在相同条件下，改进的 DNN-HMM 模型相比 DNN-HMM 模型有更强的建模能力。

4 结论

论文建立了改进的 DNN-HMM 声学模型，使用 TIMIT 语音数据集，通过语音识别评价指标句错误率和词错误率分析了不同 Fbank 特征滤波组对改进的 DNN-HMM 声学模型的影响，并与 DNN-HMM 在相同实验条件下进行了比较，证明了改进的 DNN-HMM 声学模型和 Fbank 参数拥有更强建模能力。论文在改进 DNN-HMM 模型实验过程中，发现模型前两层的 DBM 无向图模型可以有效去除噪音，而这也为论文后续的研究指明了一个方向。

参考文献

[1] Akira H, Kazunori I, Nobuo S. Marginalized Viterbi algorithm for hierarchical hidden Markov models[J]. Pattern Recognition, 2013, 46(12): 3452-3459.

[2] 胡政权, 曾毓敏, 宗原, 等. 说话人识别中 MFCC 参数提取的改进[J]. 计算机工程与应用, 2014, 50(7): 217-220.
Hu Zhengquan, Zeng Yuming, Zong Yuan, et al. Improvement of MFCC parameters extraction in speaker recognition[J]. Computer Engineering and Applications, 2014, 50(7): 217-220.

[3] 赵涛涛, 杨鸿武. 结合 EMD 和加权 Mel 倒谱的语音共振峰提取算法[J]. 计算机工程与应用, 2015, 51(9): 207-212.
Zhao Taotao, Yang Hongwu. Formant extraction algorithm of speech signal by combining EMD and WMCEP[J]. Computer Engineering and Applications, 2015, 51(9): 207-212.

[4] 侯一民, 周慧琼, 王政一. 深度学习在语音识别中的研究进展综述[J]. 计算机应用研究, 2017, 34(8): 2241-2246.
Hou Yimin, Zhou Huiqiong, Wang Zhengyi. Overview of speech recognition based on deep learning[J]. Application Research of Computers, 2017, 34(8): 2241-2246.

[5] 邓侃, 欧智坚. 深层神经网络语音识别自适应方法研究[J]. 计算机应用研究, 2016, 33(7): 1966-1970.
Deng Kan, Ou Zhijian. Adaptation method for deep neural network-based speech recognition[J]. Application Research of Computers, 2016, 33(7): 1966-1970.

[6] Mohamed A R, Sainath T N, Dahl G, et al. Deep belief networks using discriminative features for phone recognition[C]//IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2011, 125(3): 5060-5063.

[7] Tomaz R, Mirjam S M, Zdravko K. Large vocabulary continuous speech recognition of an inflected language using stems and endings[J]. Speech Communication, 2007, 49(6): 437-452.

[8] Wang J, Li L, Wang D, et al. Research on generalization property of time-varying Fbank-weighted MFCC for I-vector based speaker verification[C]// International Symposium on Chinese Spoken Language Processing. IEEE, 2014: 423-423.

[9] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.

[10] 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.

[11] 张劲松, 高迎明, 解焱陆. 基于 DNN 的发音偏误趋势检测[J]. 清华大学学报: 自然科学版, 2016, 56(11): 1220-1225.
Zhang Jinsong, Gao Yingming, Xie Yanlu. Mispronunciation tendency detection using deep neural networks[J]. Journal of Tsinghua University(Science and Technology), 2016, 56(11): 1220-1225.

[12] Kovacs G, Toth L, Compernolle D V, et al. Increasing the robustness of CNN acoustic models using autoregressive moving average spectrogram features and channel dropout[J]. Pattern Recognition Letters, 2017, 100(1): 44-50.

[13] Miloslavskaya V, Trifonov P. Sequential decoding of polar codes[J]. IEEE Communications Letters, 2014, 18(7): 1127-1130.

[14] Jiang Y B, Gao X N, Han P. Application of maximum entropy model in distribution of vehicle speed[J]. Advanced Materials Research, 2015, 1079-1080: 942-945.

[15] Lopes C, Perdigao F. Phone Recognition on the TIMIT Dataset[M]//Speech Technologies. InTech, 2011.

[16] Salakhutdinov R, Hinton G E. Deep Boltzmann machines[C]//Proceedings of International Conference on Artificial Intelligence and Statistics 2009. Brookline, MA, USA: Microtome Publishing, 2009: 448-455.

[17] Hu W J, Fu M J, Pan W L. Primi speech recognition based on deep neural network[C]//Proceedings of IEEE International Conference on Intelligent Systems. Washington D.C., USA: IEEE Press, 2016: 667-671.