ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 34,No. 2,Mar. 2019,pp. 281-287
DOI:10. 16337/j. 1004-9037. 2019. 02. 010

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:+86-025-84892742

电视剧语音识别中的半监督自动语音分割算法

龙艳花　茅红伟　叶　宏

(上海师范大学信息与机电工程学院,上海,200234)

摘　要: 针对具有大段连续文本标注、但无时间标签的电视剧语音提出了一种半监督自动语音分割算法。首先采用原始的标注文本构建一个有偏的语言模型,然后将该语言模型以一种半监督的方式用于电视剧语音识别中,最后利用自动语音识别的解码结果对传统的基于距离度量、模型分类以及基于音素识别的语音分割算法进行改进。在英国科幻电视剧"神秘博士"数据集合上的实验结果表明,提出的半监督自动语音分割算法能够取得明显优于传统语音分割算法的性能,不仅有效解决了电视剧语音识别中大段连续音频的自动分割问题,还能对相应的大段连续文本标注进行分段,保证分割后各语音段时间标签及其对应文本的准确性。

关键词: 语音识别;半监督;语音标注

中图分类号: TP918　　**文献标志码:** A

Semi-supervised Automatic Speech Segmentation for TV-drama Speech Recognition

Long Yanhua, Mao Hongwei, Ye Hong

(The College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai, 200234,China)

Abstract: To deal with the speech segmentation of TV-drama which has large coherent text transcriptions but no time-stamps, an automatic semi-supervised speech segmentation algorithm is proposed in the paper. Firstly, the original text transcriptions are used to build a biased language model, then the model is applied to the TV-drama speech recognition in a semi-supervised way, and finally, the resulting automatic speech decoding hypothesis are well combined with the traditional segmentation methods to improve the performances of speech segmentation. These traditional methods are usually based on the distance metric, model classification and the phone recognizers. Experimental results on the British TV-drama "Doctor Who" database demonstrate that, the proposed approach can achieve significant performance improvement over traditional baseline algorithms. Meanwhile, the proposed approach allows high quality segmentation and the associated transcription alignments for the large coherent TV-drama speech recordings.

Key words: speech recognition; semi-supervised; speech transcription

引　　言

近年来,语音识别的研究已超越了传统的声音到文字的转换,"电影、电视、网络视频"等多媒体音频内容的语音识别与检索已逐渐被重视[1-4]。同时,大数据时代移动互联网的发展正在改变

基金项目: 上海市青年科技英才扬帆计划(14YF1409300)资助项目;国家自然科学基金(61701306)资助项目。
收稿日期: 2017-06-25;**修订日期:** 2017-12-29

着传统语音识别语料的获取方式，特别是海量数据如新闻广播、电视语料等的获取变得越来越容易。如何对海量语料过滤，筛选及切分成适用于语音识别系统所需的音频段或句子等技术的深入研究显得日益迫切。本文将探索电视剧语音的分割算法，与以往大多语音分割方面的研究不同，本文研究的语音分割是指从复杂音频信号中提取出含有语音的语音片段，不需要按说话人进行分割，而传统的语音分割主要是指单声道语音的不重叠的多说话人语音段的分割，即按说话人不同进行分割。

与以往传统语音识别任务相比，对电视剧语音进行识别是一项非常有挑战性的工作。文献[5]中来自英国爱丁堡大学和剑桥大学2012年的研究成果显示，采用基于深度神经网络的大词汇量连续语音识别系统在广播语音识别任务上的词错误率约为10%，而其在电视剧语音识别任务上的词错误率却大大增加到约60%。这正是因为电视剧语音识别存在以下难点[4-6]：(1)自然的语音风格。与传统朗读风格的语音相比，电视剧语音大多为自然对话语音，说话人语气语调、情感变化、发音风格和习惯等变化频繁且灵活多样。(2)复杂的声学环境。不同种类背景噪声如环境噪声、动物叫声、歌声、音乐声及其各种搭配组合均有可能出现在电视剧音频中。(3)模糊的语句边界。电视剧语音内容是连续输入的，而非传统语音识别中以句子为单位的录音。通常一集电视剧都会持续数十分钟。以上难点不仅增加了大段连续电视剧语音分割的难度，同时也大大降低了语音识别的性能。

传统主流的语音分割算法可分为基于距离、模型和音素识别3大类[1,6-7]。虽然这些算法在平稳噪声环境下的连续语音识别任务中已取得较好效果，但将其直接应用到电视剧语音分割中效果不佳[6]。因此，为适应复杂环境下语音分割的需要，各种改进算法不断涌现，如文献[6]中采用的深度神经网络架构，文献[8]中基于条件随机场的多层特征分割算法等。而本文从半监督学习的角度提出了一种全新的自动语音分割算法。首先采用传统分割算法对电视剧语音进行初始分割，然后将电视剧文本标注和初始分割语音段对应的解码文本进行对齐，找到它们之间的互补性来重新更新初始分割边界。

1 传统的自动语音分割算法

1.1 传统语音分割算法

基于距离的语音分割算法主要利用相邻窗样本间的距离来度量相邻语音段的相似性，其距离度量方法有一般似然比(Generalized likelihood ratio, GLR)[9]，Kullback-Leibler(KL)距离[10]等。该类算法实现较简单，但易受语音窗长、输入特征和度量准则等影响，易检出过多的冗余分割点，且对说话人的改变较敏感。基于模型的分割算法主要以基于隐马尔科夫[11](Hidden Markov model, HMM)和高斯混合模型(Gaussian mixture model, GMM)算法为主[1]，该类算法不仅能用来对长语音段进行有效切分，而且还能对不同音频信息、说话人等进行有效聚类或分类，用以提高语音识别模型自适应能力。而基于音素识别的分割算法本质上隶属于基于HMM模型的分割算法，但由于该算法近年来在自动语音分割系统中的应用广泛且收效甚好[12]，故本文将其单独列出作为一类，该算法的实现较前两种复杂，但其对语音和非语音段的检测更加准确和精细。

1.2 自动语音分割基线系统

为了充分挖掘不同分割算法的优点，文献[7]巧妙地把基于KL距离、混合高斯模型和音素识别的三种算法结合在一起。本文将基于该算法构建的语音分割系统作为对电视剧语音进行切分的基线系统，与提出的半监督自动语音分割算法进行比较。不同的是：(1)因电视剧语音通常含大量除音乐以外的背景噪声，故本文在文献[7]的基础上引入背景噪声GMM分类器。(2)本文实验所用的语音几乎都是宽带信号，故无需考虑对语音信号进行细致的带宽划分。

基线系统架构如图1所示，首先对电视剧音频提取感知线性预测系数(Perception linear predic-

tion ,PLP)和 Mel频率倒谱系数(Mel frequency cepstrum coefficient,MFCC)。然后使用已训练好的纯语音、音乐和背景噪声GMM分类器将电视剧音频分割成纯语音、纯音乐、纯背景噪声、带噪的语音和带音乐的语音5类,其中被检测出为纯音乐和纯背景噪声的音频段在该阶段被直接丢弃。

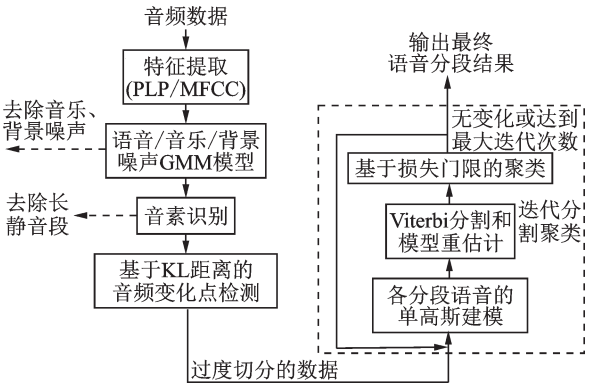

图1 自动语音分割基线系统

Fig.1 Baseline system of automatic speech segmentation

经上述粗分类后的结果接着作为音素识别器的输入进行语音和非语音的精细检测,其中音素识别器由46个上下文无关的单音素HMM模型和1个静音模型构成。46个单音素源于采用的Combilex词典(为语音识别任务稍作改动),网址:http://www.cstr.ed.ac.uk/research/projects/combilex/。连续静音大于1 s的部分被舍弃,剩下片段构成新语音段。在每一段中,采用对称KL距离d_{SD}来检测段内潜在的音频变化点[10],则

$$\begin{aligned} d_{\mathrm{SD}} &= d_{\mathrm{AHS}} + \mathrm{tr}\left[(\boldsymbol{\Sigma}_1^{-1}+\boldsymbol{\Sigma}_2^{-1})(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^{\mathrm{T}}\right] \\ d_{\mathrm{AHS}} &= \mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1})+\mathrm{tr}(\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1})-2D \end{aligned} \tag{1}$$

式中:D为特征向量维数,tr($\boldsymbol{x}$)是$\boldsymbol{x}$的秩,$\boldsymbol{\mu}$,$\boldsymbol{\Sigma}$为均值向量和协方差矩阵。

最后将上述分段结果中过度切分的片段用迭代分割聚类方法进行合并[13]:首先对每段语音进行单高斯建模,然后将这些模型用作进一步的Viterbi分割及高斯模型重估计。用来决定相邻两小段是否合并的对数似然损失函数d定义如式(2)所示[14]。那些d小于门限的小段被直接合并得到新的分割结果并重新建立新单高斯模型。该过程不断重复直到分割结果不再改变或达到最大设定的迭代次数时终止。

$$\begin{aligned} d = &\frac{N_1}{2}\log\left|\boldsymbol{\Sigma}_1\right| + \frac{N_2}{2}\log\left|\boldsymbol{\Sigma}_2\right| - \frac{N}{2}\log\left|\boldsymbol{W}\right| - \\ &\frac{N}{2}\log\left(1+\frac{N_1N_2}{N_2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\right)^{\mathrm{T}}\boldsymbol{W}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\Bigg) \end{aligned} \tag{2}$$

式中:N_1,N_2为相邻两小段的帧数,$\boldsymbol{\mu}$,$\boldsymbol{\Sigma}$为均值向量和协方差矩阵,$N=N_1+N_2$,$\boldsymbol{W}=\dfrac{N_1\boldsymbol{\Sigma}_1}{N}+\dfrac{N_2\boldsymbol{\Sigma}_2}{N}$。

2 半监督自动语音分割算法

通常语音识别系统训练语料中的人工标注信息已包含了准确的语音段起始和结束时间标签,且语音段与其对应的文本标注是对齐的,即文本标注也是事先分好段的。然而,对电视剧语料而言,往往只能获得整段电视音频及对应的大段文本标注(主要指电视剧剧本),无法得知语音段对应的具体文本段。另外,电视剧语音声学环境复杂,现有分割算法很难达到较高的准确度。因此,怎样最大限度地挖掘和利用所获得的语料信息来辅助和提升现有电视剧语音分段及其标注对齐尤为重要。

本节提出一种半监督自动语音分割算法,首先采用1.2节中的基线系统对电视剧音频进行初始分割,然后利用其半监督自动语音识别解码结果及可获得的原始文本标注对初始分割结果进行错误检测用以指导原始音频数据的重新分割,期望在达到提升电视剧语音自动分割性能的同时,还能保证分割后各语音段文本标注的自动对齐。其算法结构如图2所示,实现步骤如下:

(1)初始分割:对于输入的任意一大段电视剧音频,采用自动语音分割基线系统进行分割,得到初始

分割结果，记为SI，SI＝$\{SI_t|t=1,2,\cdots,T\}$，T为分割段数。

(2) 半监督语音识别：对SI进行自动语音识别，结果记为SD，SD＝$\{SD_t|t=1,2,\cdots,T\}$。与通常语音识别不同的是，为了最大限度提升SD准确性，全部可获得的标注文本被用来构建集内语言模型，并与采用集外文本数据训练的通用背景语言模型进行插值得到一个有偏的语言模型用于语音识别，插值权重分别为0.9和0.1。因采用了被识别数据的文本标注信息来辅助识别，故称为半监督自动语音识别[15]。

(3) DP对齐：采用动态规划(Dynamic programming，DP)对SD和原始文本标注进行对齐和比较，计算两者之间的词匹配率(Word matching rate，WMR)，WMR的计算以各段SD为参考，方法同词错误率(Word error rate，WER)，对于任意初始分割段t而言，其词匹配率记为WMR_t。

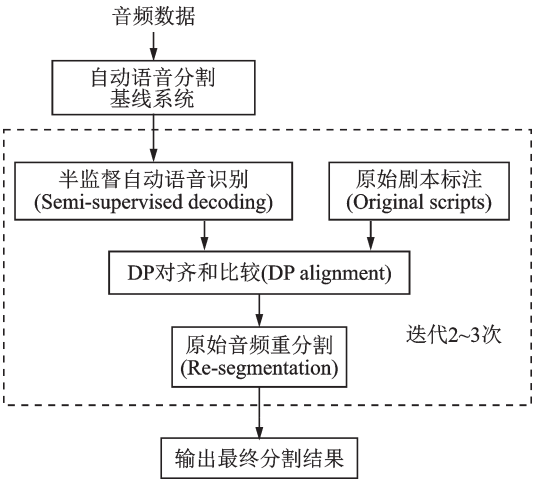

图2 半监督自动语音分割算法架构

Fig. 2 Framework of semi-supervised speech segmentation

(4) 原始音频重分割：对每一段原始音频进行如下步骤的重分割：

(a) 对SI中任意分段SI_t，若其对应的$WMR_t \leqslant TR_1$，则认为该段初始分割正确，将其从SI中移出放入新集合中，该集合记为保留段集合R。

(b) 对于SI中剩余的分割段，比较各段的SD和采用DP对齐的原始文本标注，若两者在某段起始和结束边界处相同且在该分段中间出现单词不同的持续时间≤TH秒，则将其从SI中移出放入R中。

(c) 对于经步骤(a)，(b)处理后SI中的剩余段，按段起始时间从小到大排列，比较所有分段的SD和采用DP对齐的原始文本标注。若两者在时间上连续的分段起始、或结束边界处相同，或连续相同的词序列之间的时间间隔≤TH秒，则将这些原始分割段从SI中移出合并成新的分段放入R中。

(d) 经步骤(c)处理后SI中的剩余段，从段起始时间最小的分割段开始，以当前段起始边界为起点，下一个满足以上3种情况中任意情况的起始边界为结束点之间的所有SI段合并成一个新段。若在步骤(3)中DP对齐时存在与该新段对齐的原始文本标注，则将其放入R中，否则认为其为非语音，直接舍弃。

(e) 对R中所有分割段，用原始文本标注进行强制对齐(Forced-alignment)，并根据对齐结果中连续静音帧长度情况对段边界作调整以保证各段长度≤30 s且各段边界处静音长度位于[0.06，0.50] s区间内。

(5) 声学模型更新：将步骤(e)中词匹配率≤TR_2的段挑出来，添加到图2半监督自动语音识别的训练语料中用以更新声学模型，以进一步提升SD识别率和原始音频重分割的准确率。

(6) 重复步骤(2～5)2～3次，得到最终分割好的电视剧语音段及其对应的文本标注结果。

从以上实现过程中可看出，本文所提算法借用半监督自动语音识别对传统语音分割算法进行改进，在实现对电视剧语音段时间上的切分的同时，还对大段连续的电视剧文本标注也做了分段，保证了分割后各语音段及其对应文本标注的准确性。识别是为了更好的分割，分割后的语音段又能用于建立更好的语音识别系统。对电视剧语音识别而言，大多数情况下只能获得大段的剧本文件，而无法得知剧本文件中的文本内容与语音段之间的对应关系，此时本文所提算法较传统语音分割算法的优势更加突出。

3 实验配置和结果

3.1 数据库

采用从2010年4月—2011年12月英国科幻电视剧“神秘博士-Doctor Who”剧集中挑选的16集共15.8 h,格式为MP3,采样率48 kHZ的音频作为实验数据。每一集持续时长40～48 min不等。该数据集被分成训练和测试两部分:1.4 h用作测试,14.4 h用作模型训练。测试数据采用人工借助LDC-XTrans (http://www.ldc.upenn.edu/tools/XTrans/)进行手工分割并对原始文本标注进行检查和确认,得到准确分割答案用于性能评测。自动语音分割基线系统所用数据与文献[7]一致。另外,15 h的WSJ0及Switchboard (h5train00)中随机选择的60 h数据[16]被用来训练半监督自动语音识别系统。

3.2 系统描述

系统的特征参数均采用帧长为25 ms,帧移为10 ms的39维PLP参数(12维静态PLP加上0阶导谱系数C_0,以及它们的一二阶差分)。倒谱均值和方差归一化,异方差线性判别分析(Heteroscedastic linear discriminant analysis,HLDA)被用来提升特征参数的鲁棒性。决策树聚类后的上下文相关三元音素(Tri-phone)物理状态总数为6 000, 平均每状态16个高斯分量的基于最小音素错误(Minimum phone error, MPE)训练准则的系统是图2中半监督自动语音识别系统的基本配置,初始系统训练数据为75 h。半监督自动语音分割算法中的TR_1=40%,TR_2=0,TH=3。

3.3 结果及分析

表1和表2从两种不同的角度对本文所提的语音分割算法进行性能对比。表1给出了分割算法在1.4 h测试集上的性能。作为比较,表1同时给出了准确的人工标注结果(比对时对应到每段音乐、背景噪声及语音段的起始和结束时间边界,并不是单纯的数据量统计)。可见,半监督自动语音分割算法明显优于基于KL距离、GMM和音素识别3种基础算法上的语音分割基线系统。由于“神秘博士”剧集中存在大量混合着音乐和强背景噪声的语音段,因而传统分割算法很难将其与纯语音或纯背景噪声区分开来,而本文算法由于利用了原始标注文本信息来辅助分割,从而体现出较基线系统的优势。但与人工分割标准答案相比,本文所提出的半监督自动语音分割算法仍然丢弃了6.84%的语音段(0.08 h),主要是由于某些边界模糊且背景环境复杂的语音段进行强制对齐原始文本标注时失败所致,这也是本文所提算法仍需改进的地方。

表1 分割算法在1.4 h测试集上的性能

Tab. 1 Performance of speech segmentation algorithms on 1.4 h test set h

分割算法	纯音乐	语音段	纯背景噪
基线系统	0.41	0.86	0.13
半监督自动语音分割	0.20	1.09	0.11
人工标注	0.18	1.17	0.05

注:语音段包括带背景噪声、音乐的语音段,静音包含在纯背景噪声中。

另外,通过对比人工对齐后的1.4 h原始文本标注与半监督自动语音分割算法最终输出的语音段对齐的文本标注,发现除了少数因Forced-alignment失败导致的语音段丢弃情况外,存活下来的语音段对应的原始标注与人工标注分割基本一致。

表2给出了两种自动分割算法在训练集上最终分割出的语音段时长,以及能间接反映分割段边界准确性的Forced-alignment存活率。与基线系统结果相比,本文所提算法不仅降低了语音段的丢弃率,同时在语音段上用原始剧本标注进行Forced-alignment的存活率也直接提高了6.13%,这说明半监督自动分割出来的分段边界更合理,句子完整性更好。

表3给出了采用表2中半监督自动分割出的约11 h训练语料上搭建的识别系统在1.4 h测试集上用

不同算法分割出的语音段识别性能。该识别系统的声学模型是对半监督自动语音识别系统进行区分性自适应训练得到的。从表3中可见,本文提出的语音分割算法较基线系统的词错误率相对下降了9.79%。而从分析替换、删除和插入错误容易看出,与音频分割时语音丢弃程度紧密相关的删除错误率的下降在其中起到了最主要的作用。同时,比较表3中最后两行的结果也可看出,两者识别性能相对差距的5.9%主要也来源于删除错误。因此如何尽可能多地保留电视剧音频分割时含有语音的音频段是本文有待改进之处。

表2 分割算法在14.4 h训练集上的性能

Tab. 2 Performance of speech segmentation algorithms on 14.4 h training set

分割算法	语音段/h	Forced-alignment存活率/%
基线系统	9.06	92.47
半监督自动语音分割	11.49	98.60
人工标注		

注:Forced-alignment存活率 = 采用原始剧本标注对语音段进行Forced-alignment存活下来的数据量/语音段的总长度。

表3 Doctor Who的自动语音识别性能

Tab.3 Automatic speech recognition performances on Doctor Who test set %

分割算法	替换错误率	删除错误率	插入错误率	WER
基线系统	34.11	21.27	6.12	61.50
半监督自动语音分割	33.05	15.54	6.89	55.48
人工标注	32.19	12.07	7.94	52.20

本文基于人工分割的“神秘博士”语音识别的WER=52.2%,该性能与传统安静环境下的语音识别任务性能相差甚远,这表明针对电视剧语音的大词汇量自动连续语音识别系统还有很大的性能提升空间,从表3中大于30%的替换错误和文献[5]中对于电视剧语音的研究结果中也能得出相似结论。

4 结束语

本文研究了电视剧语音识别系统中集分段和分类于一体的自动语音分割算法。文中提出的半监督自动语音分割算法能有效地将原始语音的半监督自动语音识别结果及其标注信息应用到语音分割中。分别从语音段分割准确性、强制对齐的存活率以及在采用所提分割算法得到的训练语料基础上训练的语音识别系统性能3个方面证明了文中所提出的算法明显优于传统分割算法。

由于该算法本质上是对传统基线系统分割结果进行错误检测并加以调整和修正,因而具有较好的稳定性,但在运算量上较传统算法有所增加。对“神秘博士”电视剧语音的测试表明,本文提出的算法不仅较好地解决了长音频段电视剧语音的自动分割问题,还有效地解决了原始剧本文本的分段标注对齐问题。最后,本文所提方法需要用到原始文本标注,还无法解决无任何原始文本标注的电视剧语音自动分割问题。实际上,很多情况下仅能获得具有极少量正确标注的文本信息甚至仅仅是网友提供的含大量错误标注的字幕信息,而且现在大多语音识别系统采用深度学习方法进行训练[17],如何有效地在深度学习框架下解决这类问题是笔者下一步的研究内容。

参 考 文 献:

[1] Misra A. Speech / Nospeech segmentation in web videos[C]//13th Conference in the Annual Series of INTERSPEECH Events. Portland, USA: ISCA, 2012: 1977-1980.

[2] Weng C, Juang B H, Povey D. Discriminative training using non-uniform criteria for keyword spotting on spontaneous speech [C]//13th Conference in the Annual Series of Interspeech Events. Portland, USA: ISCA, 2012: 559-562.

[3] Bell P, Swietojanski P, Renals S. Multi-level adaptive networks in tandem and hybrid ASR systems[C]//IEEE International Conference on Acoustics, Speech, and Signal Processing.Vancouver, Canada: IEEE, 2013: 6744-6748.

[4] Ryant N, Liberman M, Yuan J. Speech activity detection on YouTube using deep neural networks[C]//14th Conference in the Annual Series of Interspeech Events. Lyon, France: ISCA, 2013: 728-731.

[5] Bell P, Gales M, Lanchantin P, et al. Transcription of multi-genre media archives using out-of-domain data[C]//4th IEEE Workshop on Spoken Language Technology. Miami, USA: IEEE, 2012: 324-329.

[6] Thadani K, Biadsy F, Bikel D. On-the-fly topic adaptation for YouTube video transcription[C]//13th Conference in the Annual Series of Interspeech Events. Portland, USA: ISCA, 2012: 210-213.

[7] Tranter S E, Gales M J F, Sinha R, et al. The development of the cambridge university RT-04 diarisation system[C]//2004 Rich Transcription Workshop (RT-04). Palisades, USA:[s.n.], 2004.

[8] Saito A, Nankaku Y, Lee A, et al. Voice activity detection based on conditional random fields using multiple features[C]//11th Conference in the Annual Series of Interspeech Events. Makuhari, Japan: ISCA, 2010: 2086-2089.

[9] Gish H, Schmidt N. Text-independent speaker identification [J]. IEEE Signal Processing Magazine, 1994, 11(4): 18-32.

[10] Moreno P J, Ho P P. A new SVM approach to speaker identification and verification using probabilistic distance kernels [R]. Tech. Rep. HPL-2004-7, HP Laboratories Cambridge, 2004.

[11] Ajmera J, Mccowan I, Bourland H. Speech/music segmentation using entropy and dynamism features in a HMM classification framework [J]. Speech Communication, 2003, 40(3): 351-363.

[12] Gales M J F, Watanabe S, FoslerLussier E. Structured discriminative models for speech recognition: An overview [J]. IEEE Signal Proceeding Magazine, 2012, 29(6): 70-81.

[13] Gauvain J L, Lamel L, Adda G. The LIMSI broadcast news transcription system [J]. Speech Communication, 2002, 37(1/2): 89-108.

[14] Gish H, Siu M H, Rohlick R. Segregation of speakers for speech recognition and speaker identification [C]// IEEE International Conference on Acoustics, Speech, and Signal Processing. Vancouver, Toronto, Canada: IEEE, 1991: 873-876.

[15] Braunschweiler N, Gales M J F, Buchholz S. Lightly supervised recognition for automatic alignment of large coherent speech recordings [C]//11th Conference in the Annual Series of INTERSPEECH Events. Makuhari, Japan:ISCA, 2010: 2222-2225.

[16] Woodland P C, Povey D. Large scale discriminative training of hidden Markov models for speech recognition [J]. Computer Speech and Language, 2002, 16(1): 25-47.

[17] 麦麦提艾力·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].数据采集与处理, 2015, 30(2):365-371.
Maimaitiaili Tuerxun, Dai Lirong. Deep neural network based uyghur large vocabulary continuous speech recognition[J]. Journal of Data Acquisition and Processing, 2015,30(2): 365-371.

作者简介:

龙艳花(1983-),**通信作者**,女,博士,副研究员,研究方向:智能语音信息处理,E-mail:yanhua@shnu.edu.cn。

茅红伟(1964-),男,副教授,研究方向:信号与信息处理。

叶宏(1963-),男,副教授,研究方向:信号与信息处理。

(编辑:陈珺)