

语音识别技术分析与应用

北方民族大学电气信息工程学院 樊海花

【摘要】语音识别作为人机交互的重要方式，正在被人们所研究。针对语音识别的算法有很多，本文对基于HMM的语音识别声学建模、基于深度神经网络的声学建模、基于统计模型的语音增强算法等算法做了阐述，为对语音识别作进一步的研究做了基础。

【关键词】语音识别；HMM；深度神经网络；统计模型

1 引言

随着移动互联网的飞速发展，语音识别作为简单快捷的人机交互方式，被人们所认识，越来越多的研究被投入其中。随着各种电子设备如智能手机、家用电器成为人们生活所必不可少的组成部分，人机交互变得越来越重要，而语音识别作为最方便的人机交互方式，被人们所使用。

2 语音识别技术分析

语音识别技术的应用具体可以从贝尔实验室首先发明的英文数字识别器开始^[1]，到线性预测编码技术和动态时间规整技术应用于大规模连续词语音识别，到普林斯顿大学将隐马模型应用到语音识别过程中，再到对于解决海量语音信息识别的基于深度学习的语音识别技术的应用^[1]。语音识别本质上可以看做是一个序列分类的问题，声学建模可以看做是解决语音特征观察序列和状态序列匹配的问题。

3 语音识别技术的应用

3.1 基于HMM声学建模

隐马尔科夫模型是由隐马尔科夫链演变而来的，用于描述随机过程中的统计特性的一种方法，它的基本思想是通过一串观测值用前验概率计算公式估算出模型参数，在语音识别训练阶段，训练的数据越多，识别结果就会越接近于实际值。马尔科夫链由初始状态、转移矩阵、输出状态组成^[2]，该模型的不同已知条件和不同未知条件的组合，会产生不同的问题。在利用隐马尔科夫模型解决语音识别问题过程中，分别用到解决状态序列解码问题和解决模型参数估计问题，在语料库训练阶段，用到班姆维奇算法，推测出模型中的参数。在语音的识别阶段，采用威特比算法，当给定观测值序列和模型参数的基础上，确定最佳意义的状态序列。该模型的缺点是当隐马模型较为复杂的时候，由它训练的参数容易收敛于局部最小值。

3.2 基于深度神经网络的声学建模

基于深度神经网络的建模，可以看做是基于高斯混合模型的复杂化，高斯混合模型本质上可以看做包含一层隐含层的神经网络，隐含层节点对应各个高斯分量，输出层为HMM输出状态^[3]。DNN则是包含多个隐含层的多层网络，输入语音信号经过分解到达第一层隐含层，从声学特征空间到达隐含层所构造的新的特征空间，各个隐含层又相当于下一个隐含层的输入层，在最后一层通过softmax网络后到状态空间。DNN相当于通过增加监督信息来调谐网络。DNN模型用于估算HMM状态的后验概率，网络采用误差反向传播算法，目标函数采用交叉熵算法表示，通过求取最优的目标函数来完成训练。DNN在很长的时间里，被人们当做研究使用的热点，与此同时，问题也随之而来，在对它的多流特征融合能力的研究中，发现其在全局有监督的精细调整阶段是比较费时的。人工神经网络需要按照随机梯下降的方法，通过误差反向传播算法对训练语音库进行训练。为了解决误差反向传播算法瘦脸相对缓慢的问题，提出了多GPU对DNN进行并行训练。

3.3 基于统计模型的语音增强算法

假设语音信号和语音信号所在环境的噪声信号的傅里叶变换系数都服从统一分布，这类算法估计的是语音信号幅度谱，称为基于统计模型的语音增强算法。隐马模型还可用于检测语音停顿，对于平稳噪声，在整个语音发音段，可以使用EM算法计算噪声的最大似然估计，但自适应增益估计只有在下一语音停顿处才能获取到，这种语音增强算法在非平稳噪声环境下性能会下降非常快。基于码书的语音增强算法可以很好地克服这个问题，它引入了瞬时增益计算，在每一帧语音中，它利用了训练语音信号、噪声信号的线性预测系数和带噪声语音观测序列来计算增益函数。在基于码书的语音增强算法中，码书只提供线性预测的参数模型，它可以很好的适应非平稳噪声环境。

3.4 基于稀疏表示的语音增强技术

在白噪声的背景下，提出基于功率谱稀疏表示的语音增强算法。该研究方法用非负限制的接近于K奇异值分解的算法训练不带噪声的纯净语音的功率谱字典，采用LARS算法得到功率谱的稀疏表示，这种算法的终止条件通过控制噪声的水平，到噪声的语音信号的功率谱和用稀疏功率谱重构的功率谱之差的范数小于规定的值，就停止算法。由稀疏表示和经过训练的字典可以得到纯净语音的功率谱估计，后结合SSB-STSA方法可以完成对信号的识别。

3.5 用于视觉和视听语音识别的级联灰色立体视觉特征提取方法

尽管立体声信息最近已经广泛应用于计算机视觉任务，但立体视觉信息在视听语音识别（AVSR）系统中的并入以及是否可以提高语音准确性仍然是一个很大程度上尚未开发的领域。该领域需要解决的三个基本问题：1）立体声功能是否有利于视觉和视听语音识别？2）如果是这样，立体声功能嵌入了多少信息？3）如何在紧凑的特征向量中对平面和立体声信息进行编码？基于嵌入平面和立体声特征的不同信息，一种新的级联混合外观视觉特征（CHAVF）提取方案被提出，将平面和立体视觉信息成功地结合到一个紧凑的特征向量中，并且在视觉上评估这一新颖特征和视听连接数字识别和孤立短语识别。结果表明，立体声信息能够显著提升语音识别能力，提出的视觉特征的性能优于视觉和视听语音识别任务中其他常用的外观视觉特征。

3.6 使用径向基函数神经网络的释义识别研究

抽象是使用替代单词对给定文本的重述。释义的识别在问答，信息提取和多文档摘要等应用中至关重要。文字的词汇，句法和语义特征可以单独使用，也可以组合使用以识别释义。支持向量机（SVM），最近邻法和决策树等机器学习分类器已经被用于释义识别，SVM识别器是最受欢迎的。径向基函数神经网络（RBFNN）被设计并实现用于识别释义。该研究在微软研究释义语料库上进行了。从实验结果可以看出，RBFNN识别器在准确性方面始终优于SVM识别器，当使用词汇，句法和语义特征的组合时，实现了最佳性能。

4 总结与展望

本文以语音特征提取和声学建模为切入点，系统的阐述了几种算法在语音识别方面的应用，为正在研究语音识别技术的学者提供了一定的理论基础。随着大数据越来越深入人们的生活，语音识别必然会成为实现人机交互、推动人工智能的关键技术，海量的语音数据的累积，对人脑感知机制的不断探索与研究，未来必定会有越来越接近人脑识别语音过程的算法，语音识别技术也会越来越趋于成熟。

5 结语

近几年，虽然语音识别技术已经逐渐应用于人们的日常生活中：智能家电、服务机器人、车载系统等，但用户的语音识别率仍然不尽如人意，所以研究出识别效率高且精度高的算法应用于语音识别显得尤其重要。

参考文献

- [1]于俊婷,刘伍颖,易绵竹,李雪,李娜.国内语音识别研究综述[J].计算机光盘软件与应用,2014,10:76-78.
- [2]于大海,孙建民.浅谈语音识别技术的应用和发展[J].科技传播,2009,09:22-23.
- [3]吴大为.基于HMM模型改进算法的语音识别系统[D].哈尔滨工业大学,2013.
- [4]周盼.基于深层神经网络的语音识别声学建模研究[D].中国科学技术大学,2014.

作者简介:

樊海花（1991—），女，山西原平人，硕士研究生。