学校编码：10384　　分类号_____密级_____
学　　号：200340016　　UDC _____

厦门大学

硕 士 学 位 论 文

改进的 HMM 与 BP 神经网络混合模型在语音识别中的应用研究

Application of Improved Hybrid Model Based on HMM and BP Neural Nework in Speech Recognition

息晓静

指导教师姓名：　林坤辉　副教授
专 业 名 称：　计算机应用技术
论文提交日期：　2006 年 4 月
论文答辩时间：　2006 年 6 月
学位授予日期：　2006 年　　月

答辩委员会主席：_____
评　　阅　　人：_____

2006 年 5 月

厦门大学学位论文原创性声明

兹呈交的学位论文，是本人在导师指导下独立完成的研究成果。本人在论文写作中参考的其他个人或集体的研究成果，均在文中以明确方式标明。本人依法享有和承担由此论文产生的权利和责任。

声明人（签名）：

年　　月　　日

厦门大学学位论文著作权使用声明

本人完全了解厦门大学有关保留、使用学位论文的规定。厦门大学有权保留并向国家主管部门或其指定机构送交论文的纸质版和电子版，有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆被查阅，有权将学位论文的内容编入有关数据库进行检索，有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

本学位论文属于

1、保密（　），在　　年解密后适用本授权书。

2、不保密（　）

（请在以上相应括号内打“√”）

作者签名：　　日期：　　年　月　日

导师签名：　　日期：　　年　月　日

内容摘要

语音识别是一门内容丰富、应用广泛的技术。本文着眼于汉语语音识别的主要问题，研究汉语语音孤立词识别的关键技术，以提高语音的识别率和识别模型的收敛速度。

本文论述了语音识别的基本原理，从语音信号的时域、频域、倒谱域出发，对语音信号进行分析，介绍了语音信号分析方法中的滤波器组分析方法和线性预测编码技术，并推导了线形预测倒谱系数(LPCC)和 Mel 倒谱系数(MFCC)。在特征提取中，选用了基于听觉模型的 MFCC，并与基于发声模型的 LPCC 参数进行分析比较。

隐马尔可夫模型(HMM)和人工神经网络在语音信号处理中都有广泛的应用，本文剖析了两者在语音信号处理上各自的优缺点。为取 HMM 和人工神经网络这两种模型各自的优异特性，在本文研究的语音识别模型中，采用它们的混合模型，并提出了一种新的结合方式。即，将 HMM 的最佳状态序列的输出概率作为人工神经网络的输入。一方面由于 BP 神经网络能够根据提供的数据，通过训练和学习，找出输入输出的内在关系，不需要一个明确的数学解析式；另一方面由于离散隐马尔可夫模型(DHMM)会产生量化的误差，所以采用连续密度隐马尔可夫模型（CDHMM）和反向传播（Back Propagation）神经网络相结合的方式，充分利用了 CDHMM 的时域建模和 BP 神经网络强大的分类能力，同时充分考虑了孤立词语音的类间特性。实验表明这种结合方式在一定程度上提高了语音的识别率。

本文还分析了传统 BP 网络训练上的局限性，在前人研究的基础上，对神经元采用更一般 tan-sigmoid 函数。在训练过程中，调整权值的同时对缩放系数和位移参数进行动态调整，使信息分布存储于权值矩阵及转换函数中，比传统的算法具有更强的非线性映射能力，实验表明这种改进的 BP 神经网络训练算法能够加快网络的收敛速度，而且能够在一定程度上克服传统训练算法容易收敛到局部极小值的局限性，从而提高了网络的收敛精度。

关键词：CDHMM；BP 神经网络；转换函数

Abstract

Speech recognition is a technology which has rich content and has been widely used. This thesis focuses on the main issues of Chinese speech recognition. In order to improve the recognition ratio and speed up convergence, the key technologies of the Chinese speech recognition has been researched.

This thesis analyzes the speech signal and describes the principle of speech recognition from the time domain, frequency domain and cepstrum domain. The filter banks analysis method and linear predictive coding technology are introduced, and the LPCC (Linear Predictive Coding Cepstrum) parameters and the MFCC (Mel-Frequency Cepstrum Coefficients) are given. In feature extraction, the MFCC parameters based on the human auditory model are chosen, and compared with the LPCC parameters based on the human phonation model.

Hidden Markov Model (HMM) and Artificial Neural Network (ANN) have been widely used in speech recognition. Their respective advantages and disadvantages are analyzed and the combination methods of Hidden Markov Model and Artificial Neural Network are summarized. In view of the advantages and disadvantages of the hybrid models, a new hybrid approach is put forward. In the new method, the states output probabilities of Continuous Density Hidden Markov Model (CDHMM) are used as the input of the Back Propagation (BP) neural network. On the one hand, the BP neural network does not need a clear formula, because it can find out the intrinsic relationship between the input and the output by training and learning, according to the provided data. On the other hand, Discrete Hidden Markov Model (DHMM) will produce the quantization error, so the CDHMM/BP neural network hybrid model is adopted, for it takes advantage of the time domain modeling ability of the CDHMM and the strong classification ability of the BP neural network, and considers the characters of different classification sufficiently. Experimental results show that the hybrid method can improve the recognition ratio to a certain extent.

This thesis summarizes the limitations of traditional BP training algorithm. Based on the former research, the commoner tan-sigmoid transform function is adopted. During the training, the zoom coefficients and the displacement parameters are adjusted with the weight matrix. The information is stored in weight

matrix and in transform function dispersedly. Its non-linear mapping ability is stronger than the traditional algorithm. Experimental results show that the improved training algorithm can speed up the convergence of the neural network, and at the same time, it can also avoid the premature convergence, thus increase the precision of the results.

Key Words: CDHMM; BP neural network; transform function

目录

Contents

第一章 绪论

1.1 语音识别概述

让机器“听懂”人的语言，并根据其信息执行人的意图，是最理想的人机智能接口方式。语音识别技术以语音信号为研究对象，涉及语言学，计算机科学，信号处理，生理学，心理学等诸多领域，是模式识别的重要分支，该技术有非常广阔的应用前景。

1.1.1 语音识别分类

1. 按词汇量大小分

每个语音识别系统都有一个词汇表，系统只能识别词汇表中所含的词条。按照词汇表大小来分，有小词汇表(词汇量小于 100)、中词汇表(词汇量在 100 和 1000 之间)、大词汇表(词汇量在 1000 词以上)语音识别。一般而言，随着词汇表中词汇量的增多，各词汇量之间的混淆性增加，系统的实现变得更加困难，系统的识别率也会降低。

2. 按发音方式分

语音识别按照语音的发音方式来分，可以分为孤立词识别、连接词识别、连续语音识别 3 种方式。所谓孤立词识别就是在发待识别语音时，每次只含词汇表中的一个词条。连接词识别是每次说词汇表中的若干个词条来进行识别，这些词条以慢速连读的方式说出，一般指从“0”到“9”十个汉语数字语音连接而成的多位数字的识别，并包含其它一些少量的操作指令等。连续语音识别指说话人以日常自然的方式讲述并进行识别。

3. 按说话人的限定范围分

有特定人识别和非特定人识别两种方式。所谓特定人识别是指识别系统只针对特定某个用户进行识别工作的方式；非特定人识别是指识别系统可以针对任何人工作。对于前者，只能识别特定人的声音，其他人要想使用该系统，必须事先输入大量的语音数据，对系统进行训练；而对后者，机器能识别任何人的发音。由于语音信号的可变性很大，这种系统要从大量人的发音样本

中学到特定人的发音速度、语音强度、发音方式等基本特征，并归纳出其相似性作为识别的标准。使用者无论是否参加过训练都可以共用一套参考模板进行语音识别。

4．按照识别方法分

有模板匹配法、概率模型法等。所谓的模板匹配法是指把不同内容的语音转换成不同的模板，并基于对模板匹配的相似度量进行语音识别的方法。而概率模型法主要是指利用隐马尔科夫模型的概率参数来对似然函数进行估计判决而得到识别结果的方法。除了以上方法外，另外还包括基于人工神经网络、支持向量机等方法的语音识别技术。

1.1.2 语音识别单元

选择识别单元是语音识别的第一步，语音识别单元有单词(句)、音节和音素三种。

单词(句)单元广泛应用于中小词汇语音识别系统，但不适合大词汇系统，原因在于模型库庞大，训练模型任务繁重，模型匹配算法复杂，难以满足实时性要求。

音节单元多见于汉语语音识别，主要因为汉语是单音节结构的语音，而英语是多音节，而且汉语虽然大约有 1300 个音节，但若不考虑声调，约有 408 个无调音节，数量相对较少，因此，对于中、大词汇量汉语语音识别系统来说，以音节为识别单元基本是可行的。

音素单元以前多见于英语语音识别的研究中，但目前中、大词汇量汉语语音识别系统也在越来越多地采用。原因在于汉语音节仅由声母(包括零声母有 22 个)和韵母(共有 28 个)构成，但是由于协同发音的影响，音素单元不稳定，所以如何获得稳定的音素单元，还有待研究。

1.1.3 语音识别原理

语音识别本质上是模式识别的过程，由训练和识别两个过程所组成。训练是指用一定数量的样本(训练集或学习集)进行分类器的设计。识别是指用所设计的分类器对待识别的样本进行分类决策。语音识别系统主要由 4 个部分组成：数据获取；预处理；特征提取和选择；分类决策。对应的语音识别系