

语音识别技术的研究与发展

王敏妲

(同济大学 微电子研究中心, 上海 201804)

摘要: 回顾了语音识别技术的发展历史,描述了语音识别系统的基本原理,介绍了语音识别的几种基本方法,并对语音识别技术面临的问题和发展前景进行了讨论。

关键词: 语音识别;模式识别;特征提取;人机交互

中图分类号: TP392.42

文献标识码: A

The research and progress of speech recognition technology

WANG Min Da

(Microelectronics Center of Tongji University, Shanghai, 201804, China)

Abstract: This paper reviews the development history of speech recognition technology and describes the fundament of speech recognition system, introduces several basic methods of speech recognition and discusses the difficulties and prospect of speech recognition development.

Key words: speech recognition; pattern recognition; feature extraction; human-computer interact

1 语音识别技术概述

语音识别是解决机器“听懂”人类语言的一项技术。作为智能计算机研究的主导方向和人机语音通信的关键技术,语音识别技术一直受到各国科学界的广泛关注。如今,随着语音识别技术研究的突破,其对计算机发展和社会生活的重要性日益凸现出来。以语音识别技术开发出的产品应用领域非常广泛,如声控电话交换、信息网络查询、家庭服务、宾馆服务、医疗服务、银行服务、工业控制、语音通信系统等,几乎深入到社会的每个行业和每个方面。

广泛意义上的语音识别按照任务的不同可以分为4个方向:说话人识别、关键词检出、语言辨识和语音识别^[1]。说话人识别技术是以语音对说话人进行区别,从而进行身份鉴别和认证的技术。关键词检出技术应用于一些具有特定要求的场合,只关注那些包含特定词的句子,例如对一些特殊人名、地名的电话监听等。语言辨识技术是通过分析处理一个语音片断以判别其所属语言种类的技术,本质上也是语音识别技术的一个方面。语音识别就是通常人们所说的以说话的内容作为识别对象的技术,它是4个方面中最重要和研究最广泛的一个方向,也是本文讨论的主要内容。

2 语音识别的研究历史及现状

语音识别的研究工作始于20世纪50年代,1952年Bell实验室开发的Audry系统是第一个可以识别10个英文数字的语音识别系统。1959年,Rorgie和Forge采用数字计算机识别英文元音和孤立词,从此开始了计算机语音识别。60年代,苏联的Matin等提出了语音结束点的端点检测,使语音识别水平明显上升;Vintsyuk提出了动态编程,这一提法在以后的识别中不可或缺。60年代末、70年代初的重要成果是提出了信号线性预测编码(LPC)技术和动态时间规整(DTW)技术,有效地解决了语音信号的特征提取和不等长语音匹配问题;同时提出了矢量量化(VQ)和隐马尔可夫模型(HMM)理论。

80年代语音识别研究进一步走向深入:HMM模型和人工神经网络(ANN)在语音识别中成功应用。1988年,FULEE Kai等用VQ/I-IMM方法实现了997个词汇的非特定人连续语音识别系统SPHINX。这是世界上第1个高性能的非特定人、大词汇量、连续语音识别系统。

进入90年代后,语音识别技术进一步成熟,并开始向市场提供产品。许多发达国家如美国、日本、韩国以及IBM、Apple、AT&T、Microsoft等公司都为语音识别系统的实用化开发研究投以巨资。同时汉语语音识别也越来越

综述与评论 Review and Comment

受到重视。IBM 开发的 ViaVoice 和 Microsoft 开发的中文识别引擎都具有了相当高的汉语语音识别水平。

进入 21 世纪,随着消费类电子产品的普及,嵌入式语音处理技术发展迅速^[2]。基于语音识别芯片的嵌入式产品也越来越多,如 Sensory 公司的 RSC 系列语音识别芯片、Infineon 公司的 Unispeech 和 Unilite 语音芯片等,这些芯片在嵌入式硬件开发中得到了广泛的应用。在软件上,目前比较成功的语音识别软件有:Nuance、IBM 的 ViaVoice 和 Microsoft 的 SAPI 以及开源软件 HTK,这些软件都是面向非特定人、大词汇量的连续语音识别系统。

我国语音识别研究一直紧跟国际水平,国家也很重视。国内中科院的自动化所、声学所以及清华大学等科研机构 and 高校都在从事语音识别领域的研究和开发。国家 863 智能计算机专家组为语音识别技术研究专门立项,并取得了高水平的科研成果。我国中科院自动化所研制的非特定人、连续语音听写系统和汉语语音人机对话系统,其准确率和系统响应率均可达 90% 以上。

3 语音识别系统

语音识别本质上是一种模式识别的过程,未知语音的模式与已知语音的参考模式逐一进行比较,最佳匹配的参考模式被作为识别结果。图 1 是基于模式匹配原理的自动语音识别系统原理框图。



图 1 语音识别系统原理框图

(1) 预处理模块:对输入的原始语音信号进行处理,滤除掉其中的不重要的信息以及背景噪声,并进行语音信号的端点检测、语音分帧以及预加重等处理。

(2) 特征提取模块:负责计算语音的声学参数,并进行特征的计算,以便提取出反映信号特征的关键特征参数用于后续处理。现在较常用的特征参数有线性预测(LPC)参数、线谱对(LSP)参数、LPCC、MFCC、ASCC、感觉加权的线性预测(PLP)参数、动态差分参数和高阶信号谱类特征等^[1]。其中,Mel 频率倒谱系数(MFCC)参数因其良好的抗噪性和鲁棒性而应用广泛。

(3) 训练阶段:用户输入若干次训练语音,经过预处理和特征提取后得到特征矢量参数,建立或修改训练语音的参考模式库。

(4) 识别阶段:将输入的语音提取特征矢量参数后与参考模式库中的模式进行相似性度量比较,并结合一定的判别规则和专家知识(如构词规则,语法规则等)得出最终的识别结果。

4 语音识别的几种基本方法

当今语音识别技术的主流算法,主要有基于动态时

间规整(DTW)算法、基于非参数模型的矢量量化(VQ)方法、基于参数模型的隐马尔可夫模型(HMM)的方法、基于人工神经网络(ANN)和支持向量机等语音识别方法。

4.1 动态时间规整(DTW)

DTW 是把时间规整和距离测度计算结合起来的一种非线性规整技术,是较早的一种模式匹配和模型训练技术。该方法成功解决了语音信号特征参数序列比较长时不等长的难题,在孤立词语音识别中获得了良好性能。

4.2 矢量量化(VQ)

矢量量化是一种重要的信号压缩方法,主要适用于小词汇量、孤立词的语音识别中。其过程是:将语音信号波形的 k 个样点的每 1 帧,或有 k 个参数的每 1 参数帧,构成 k 维空间中的 1 个矢量,然后对矢量进行量化。量化时,将 k 维无限空间划分为 M 个区域边界,然后将输入矢量与这些边界进行比较,并被量化为“距离”最小的区域边界的中心矢量值。矢量量化器的设计就是从大量信号样本中训练出好的码书,从实际效果出发寻找到的失真测度定义公式,设计出最佳的矢量量化系统,用最少的搜索和计算失真的运算量,实现最大可能的平均信噪比。

4.3 隐马尔可夫模型(HMM)

隐马尔可夫模型是 20 世纪 70 年代引入语音识别理论的,它的出现使得自然语音识别系统取得了实质性的突破。目前大多数大词汇量、连续语音的非特定人语音识别系统都是基于 HMM 模型的。

HMM 是对语音信号的时间序列结构建立统计模型,将其看作一个数学上的双重随机过程:一个是用具有有限状态数的 Markov 链来模拟语音信号统计特性变化的隐含的随机过程,另一个是与 Markov 链的每一个状态相关联的观测序列的随机过程。前者通过后者表现出来,但前者的具体参数是不可测的。人的言语过程实际上就是一个双重随机过程,语音信号本身是一个可观测的时变序列,是由大脑根据语法知识和言语需要(不可观测的状态)发出的音素的参数流。HMM 合理地模仿了这一过程,很好地描述了语音信号的整体非平稳性和局部平稳性,是较为理想的一种语音模型。

HMM 模型可细分为离散隐马尔可夫模型(DHMM)和连续隐马尔可夫模型(CHMM)以及半连续隐马尔可夫模型(SCHMM)等^[3]。

4.4 人工神经网络(ANN)

人工神经网络在语音识别中的应用是目前研究的又一热点。ANN 实际上是一个超大规模非线性连续时间自适应信息处理系统,它模拟了人类神经元活动的原理,最主要的特征为连续时间非线性动力学、网络的全局作用、大规模并行分布处理及高度的稳健性和学习联想能力。这些能力是 HMM 模型不具备的。但 ANN 又不具有 HMM 模型的动态时间归正性能。因此,人们尝

综述与评论 Review and Comment

试研究基于 HMM 和 ANN 的混合模型,把两者的优点有机结合起来,从而提高整个模型的鲁棒性,这也是目前研究的一个热点。

4.5 支持向量机(SVM)

支持向量机是应用统计学学习理论的一种新的学习机模型,它采用结构风险最小化原理(SRM),有效克服了传统经验风险最小化方法的缺点,在解决小样本、非线性及高维模式识别方面有许多优越的性能^[4]。其基本思想可以概括为:首先通过非线性变换将输入空间变换到一个高维空间,然后在这个新空间中求取最优线性分类面,而这种非线性变换是通过定义适当的内积函数实现的。目前,统计学学习理论和支持向量机也是国际上机器学习领域的研究热点。

5 语音识别所面临的问题

尽管语音识别取得很大成功,但是距离真正的人机自由交流还有很大的距离。例如,目前计算机还需要对用户做大量训练才能更准确识别,用户的语音识别率也并不是尽如人意。主要难题有以下几个方面:

(1)识别系统的适应性差。主要体现在对环境依赖性强,特别在高噪音环境下语音识别性能还不理想。

(2)语音识别系统从实验室演示系统到商品的转化过程中,还有许多具体问题需要解决。例如,识别速度、拒识等问题,还有连续语音中去除不必要语气词如“呃”、“啊”等语音的技术细节问题。

(3)语言学、生理学、心理学方面的研究成果已有不少,但如何把这些知识量化、建模并用于语音识别,还需要进一步研究。

面对上面的困难,语音识别技术要做到真正成功,在任何环境中都能人机进行自由地对话,不仅需要语音识别基础理论的突破,更需要大量的实际工作的积累。

6 语音识别技术的前景展望

语音作为当前通信系统中最自然的通信媒介,语音识别技术是非常重要的交互技术。随着计算机和语音处理技术的发展,语音识别系统的实用性将进一步提高。应用语音的自动理解和翻译,可消除人类相互交往的语言障碍。国外已有多种基于语音识别产品(如声控拨号电话、语音记事本等)的应用,基于特定任务和环境的听写机也已经进入应用阶段。这预示着语音识别技术有着非常广泛的应用领域和市场前景。随着语音技术的进步和通信技术的飞速发展,语音识别技术将为网上会议、商业管理、医药卫生、教育培训等各个领域带来极大的便利^[5],其应用和社会效益前景非常良好。

参考文献

- [1] 王炳锡,屈丹,彭焱.实用语音识别基础[M].北京:国防工业出版社,2005.
- [2] 刘么和,宋庭新.语音识别与控制应用技术[M].北京:科学出版社,2008.
- [3] 詹新明,黄南山,杨灿.语音识别技术研究进展[J].现代计算机,2008,291(9):43-45.
- [4] 柳春.语音识别技术研究进展[J].甘肃科技,2008,24(9):41-43.
- [5] 高新涛,陈乖丽.语音识别技术的发展现状及应用前景[J].甘肃科技纵横,2007,36(4):13.

(收稿日期:2009-08-13)