

小波神经网络在语音识别系统中的应用

宋 弘

(四川理工学院 自动化与电子信息学院, 四川 自贡 643000)

摘 要: 文中设计了以“MCU+DSP”的双 CPU 结构为核心的语音识别系统,利用 MCU 实现语音识别系统的控制,DSP 实现语音识别系统的预处理、特征提取和模式识别.针对语音识别系统各个阶段,分别引入小波变换和神经网络对传统的实现方法进行改进,有效地降低了系统中的噪声,减少了系统中不必要的数据量,提高了系统的语音识别性能.实验数据验证了该方法的可行性和有效性.

关键词: 小波变换;神经网络;语音识别系统

中图分类号: TP274

文献标识码: A

文章编号: 1000-7180(2010)04-0122-03

Application of Wavelet Neural Network in Speech Recognition System

SONG Hong

(School of Automation and Electronic Information Engineering,
Sichuan University of Science & Engineering, Zigong 643000, China)

Abstract: A voice recognition system based on a double CPU construction with DSP and MCU is designed in the paper. The MCU is used in the control of the voice recognition system. The DSP is used for realizing the preprocessing, the feature extraction and the pattern identification of the voice recognition system. According to the different situation of each stage of the voice recognition system, the wavelet transform and the neural net are introduced into the improvement of traditional realizing methods, which will reduce noises of the system effectively, decrease unnecessary data greatly, and improve the performance of the voice recognition system. The experimental data verify the feasibility and the validity of the method proposed.

Key words: wavelet transform; neural net; voice recognition system

1 引言

人与机器之间的自然快捷而又稳定可靠的交互方式是人类追求的理想,也是科技人员孜孜追求的方向.由于传统的语音识别存在很多限制^[1],文中设计了以“MCU+DSP”的双 CPU 结构为核心的语音识别系统,充分利用数字信号处理器(DSP)超强运算能力,以及单片机(MCU)善于控制的特点.同时,利用小波变换具有良好的时频特性,能较好地解决时阈和频阈分辨率的矛盾,将其引入到语音识别系统的预处理和特征提取阶段,对传统的实现方法

进行改进,有效地降低了系统中的噪声;而神经网络具有自学习、自适应、鲁棒性、容错性和推广能力,将其引入到语音识别系统的模式识别阶段,提高了系统的语音识别性能.

2 语音识别系统的硬件结构

该系统的硬件设计采用模块化设计方案,采用不同功能模块进行相关硬件电路的设计.由于单片机具有善于控制的特点,因此语音识别系统控制由单片机来完成.由于数字信号处理器具有超强运算能力,因此语音识别系统的特征提取、训练算法和模

收稿日期: 2009-03-16; 修回日期: 2009-06-20

基金项目: 四川省教育厅重点项目(2006ZD036);人工智能四川省重点实验室重点项目(2007RZ008,2008RZ001)

式识别阶段的工作由数字信号处理器来完成,如图1所示.系统由TMS320C32 DSP芯片和C8051F021 MCU芯片作为核心,配合A/D电、D/A电路、控制电路、接口电路、键盘、存储电路来实现语音识别系统.

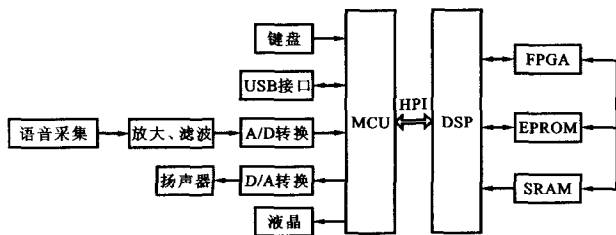


图1 语音识别系统的硬件结构框图

TI公司的TMS320C32是一款32位浮点数字信号处理芯片,其内部存储空间大,运算功能强,运行速度可达40 MFLOPS,片内RAM为 512×32 bit,16 M片外存储空间,可重新定位的中断向量表以及可选的边缘、电平触发中断方式等.该芯片的特点使得该芯片构成体积小、性能高、成本低,是一款理想的DSP处理器.同时,由于该芯片还含多路MCBSP(多路缓冲串口),在一定程度上减轻了MCU接口扩展的负担.C8051F021含有丰富的数字外设,包括4个I/O端口,可同时使用的硬件SM-BUS,SPI和2个增强性UART串口,5个通用的16位计数器/定时器,专用的看门狗定时器^[2-3].

DSP与MCU之间的快速、可靠的数据通信是系统设计成功的关键之一.文中利用DSP提供的HPI接口实现了双CPU间数据通信,数据传输速度可达160 MB/s以上.此时,DSP中的RAM相当于一个双口RAM,MCU通过HPI接口对此RAM进行操作,而DSP则是操作自己的内存^[4].

3 小波变换与神经网络

3.1 小波变换

小波变换是将信号与一个在时域和频域均具有良好局部化性质的函数族进行卷积积分,从而把信号分解成一组位于不同频带和时段内分量.

小波函数的定义为:设 $\phi(t)$ 为一平方可积函数 $\phi(t) \in L^2(R)$,若 $\bar{\phi}(\omega)$ (其傅立叶变换)满足允许条件(Admissible Condition):

$$C_{\phi} = \int_R \frac{|\bar{\phi}(\omega)|^2}{|\omega|} d\omega < \infty \quad (1)$$

则称 $\phi(t)$ 为基本小波或母小波的函数.由同一基本小波 $\phi(t)$ 经伸缩和平移后得到的一组函数称为一族小波或一个小波序列^[5-6].

小波变换的实质就是采用一族小波函数去表示信号或函数,同时,小波函数又可看成一个带通滤波器的冲激响应,经小波变换后数据将分解到一系列频带上,并可在不同频带中进行信号处理与分析.我们希望在丢失原信号信息的情况下,尽量减小小波变换系数的冗余度.常用的方法是将尺度按幕级数进行二进制离散.

3.2 神经网络

人工神经网络(Artificial Neural Network, ANN)又称神经元网络、神经网络.它是80年代后期迅速发展起来的一门新兴科学,能够模拟人脑的某些智能行为,具有对信息进行并行处理、分布式存储以及自学习与推理能力,以及易于硬件实现等优点,受到众多专业人员的关注,在诸多领域取得了成功应用.构成人工神经网络的三个基本要素是神经元、网络拓扑、网络的训练学习算法.其中BP算法是目前应用最为广泛、最活跃的神经网络学习算法^[4-5].BP神经网络是由输入层、隐含层、输出层组成,每一层的神经元只接受来自前一层的神经元的信息.通过反向传播,调整权值,使输入和输出之间具有较好的映射作用.

4 语音识别系统的预处理

语音识别系统中的预处理包括:预加重、加窗、端点检测.预处理单元的输入是采样得到的原始语音信号,输出是去除了非语音段的数帧语音信号.由于语音是不平稳的随机过程,其特征是随着时间缓慢变化的,预加重的作用是将高频提升,以弥补声音在唇部辐射时产生的高频损失;对信号加窗以避免短时语音段边缘的影响;端点检测的作用是正确区分语音信号与各种背景噪声.

语音信号在获取过程中,由于各种客观因素的影响在检测到的数据中常混有噪音,经A/D转换后又增加了量化噪音,同时,无声段或者发音前后人为呼吸等产生的杂音使得语音的端点比较模糊,因此,在提取特征参数前,要对数据进行预处理,尽可能去除这些噪音以提高信噪比^[6].文中利用小波变换的多分辨率特点将原始语音信号分解为多个帧,计算各帧的能量和方差,同时,改变小波变换系数进行降噪处理.小波变换消噪方法主要选择适当的小波函数和小波分解层数将含有噪音的原始数据输出序列进行正交小波变换,确定变换次数,通过设置阈值,对数据进行离散小波变换,计算所有小波系数,然后剔除被认为跟噪声有关的小波系数,并小波系数进

行修正,对修正后的小波系数进行逆变换,得到降噪后的数据.

5 语音识别系统的特征提取

特征提取是语音识别的一个重要环节.它既是一个信息大幅度压缩的过程,也是一个信号解卷的过程.语音识别中对特征参数的要求是:(1)能有效代表语音特征;(2)各阶参数之间有良好的独立性;(3)特征参数要计算方便,以保证语音识别的实时实现^[2].

小波分析正是一种处理随时间缓慢变化信号的有效方法.利用小波变换将降噪后的语音信号分解,然后基于一定的选择规则,得到在各个频率段的决策表,对决策表进行一致性验证得到最终的决策表,最后在最终的决策表的基础上,将每个分辨率下小波系数的频谱合成完整频谱,从而得到特征参数.

6 语音识别系统的模式识别

语音识别系统在信号的特征参数提取出来后需要进行模式识别.模式识别是指按照一定的准则,从大量已知模式中获取表征该模式本质特征的模型参数.语音识别所应用的模式识别方法主要有动态时间规整技术(DTW)、隐马尔可夫模型(HMM)和人工神经网络(ANN).本系统采用人工神经网络.人工神经网络是由大量的神经元连接而成的,是一种大规模、分布式的神经元处理系统.由于模式识别过程接近人类思维活动,与人脑神经系统有较强的相似性,因此利用神经网络的结构优势和高速的并行运算能力进行模式识别是一种有效的技术途径.

文中所构建的神经网络采用 BP 神经网络如图 2 所示.它是由输入层、隐含层、输出层和非线性传递函数组成的三层感知器网络.其学习过程由信号的正向传播与误差的反向传播两个过程组成.输入向量为 $X = (x_1, \dots, x_n)^T$,各输入值为语音信号的特征参数;输出层输出向量为 $Y = (y_1, \dots, y_l)^T$;隐含层节点数的选取是根据经验公式 $m = \sqrt{nl}$ 确定的,因此隐含层向量为 $O = (o_1, \dots, o_m)^T$.输入层到隐含层之间的权值矩阵用 V 表示, $V = (V_1, \dots, V_m)$;隐含层到输出层之间的权值矩阵用 W 表示, $W = (W_1, W_2, \dots, W_l)$.

对于输入层有:

$$o_k = f(\text{net}_k) = f\left(\sum_{j=0}^5 w_{jk}x_j\right) \quad (2)$$

$$k = 1, 2, \dots, l$$

对于隐含层有:

$$y_j = f(\text{net}_j) = f\left(\sum_{i=0}^9 v_{ij}x_i\right) \quad (3)$$

$$j = 1, 2, \dots, m$$

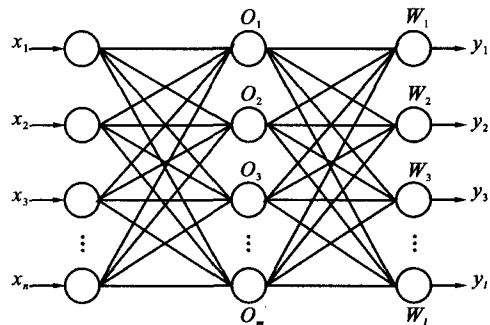


图2 三层 BP 神经网络

正向传播时,模式作用于输入层,经隐含层处理后,传向输出层.如输出层未能满足预期的输出要求,则转入误差的反向传播阶段,将输出误差按某种子形式,通过隐含层向输入层逐层返回,并“分摊”给各层的所有单元.从而获得各层单元的参考误差或称误差信号,以作为修改各单元权值的依据.这种信号正向传播与误差反向传播的各层权值矩阵的修改过程,是周而复始进行的.权值不断修改的过程,也就是网络不断学习(训练)的过程.此过程一直进行到网络输出的误差逐渐减少到可接受的程度或达到设定的学习次数为止.

7 实验结果分析

在实验室条件下录制了“开门”、“关门”、“停止”、“坐下”、“起立”、“前进”等 50 句常用语音命令,参与训练者分别是 10 个男生和 10 个女生.对语音信号采用 16kHz 进行采样.训练时取前五遍来建立这句语音命令的训练模版,其余作为输入来检测识别结果.语音识别结果的正确率为 97.5%.实验表明基于小波变换预处理和特征提取,并采用人工神经网络进行模式识别后,有效降低了系统中的噪声,减少了系统中不必要的数据量,使得语音识别更为有效.

8 结束语

文中采用了以“MCU + DSP”的双 CPU 结构为核心的语音识别系统,充分利用单片机(MCU)善于控制的特点进行语音识别系统控制;充分利用数字信号处理器(DSP)超强运算能力,实现语音识别系

(下转第 128 页)

(6) 使启动信号 $CONVST = 0$, 如果继续转换其他通道的模拟量, 则转(2), 否则转(1);

根据以上操作步骤, 每采集完一个通道, 则 FPGA 对 AD7890 的控制寄存器重新初始化, 指向下一通道, 如此循环往复, 完成多路模拟量的分时采集. 硬件测试实验表明, 此方法模拟量采集速度快, 性能稳定.

4 系统验证

本设计在 Xilinx ISE^[7] 和 Modelsim 的软件平台上进行了综合以及 Post-map 仿真, 并在实际工程中得到了验证. 利用周立功公司的 USBCAN-II 设备产生并发送 CAN 总线数据, 在模拟量输入端施加不超过 $\pm 10V$ 的模拟量, 进行 FPGA 片上数据采集测试实验. 用 Xilinx 公司的 ChipScope 软件做片上数据测试. 由实验结果可知, 该装置完全满足 CAN 总线数据通信的要求, 实现了多通道的模拟量采集.

5 结束语

文中提出的基于 FPGA 的适用于小卫星地检的数据采集及通信装置, 既适用于地面测试设备的研制, 又为航天领域小卫星的星上数据采集提供了不同于单片机控制的一种可以实现的方案. 该方案在解决一些可应用在航天领域的单片机速度较低, 资源较少, 功耗较高, 容易出现程序跑飞等问题方面

提供了新的途径.

实验结果证明本设计方法正确、使用方便、可扩展性强、运行稳定可靠, 对于 CAN 总线在实际工程的应用有一定的参考价值和应用前景.

参考文献:

- [1] 黄志澄. 小卫星技术及其军事应用的新进展[J]. 卫星应用, 2008, 16(2): 36-41.
- [2] 周新发, 尚志, 刘群. 工业现场 CAN 总线在航天领域的应用[J]. 现场总线与网络技术, 2006(1): 60-63.
- [3] 刘涛. CAN 总线接口电路设计中的关键问题[J]. 工矿自动化, 2007(1): 100-102.
- [4] 汤楠, 李明. AD7890 型串行数/模转换器与 51 系列单片机的接口设计[J]. 国外电子元器件, 2006(8): 4-7.
- [5] 饶运涛, 邹继军. 现场总线 CAN 原理与应用技术[M]. 2 版. 北京: 北京航空航天大学出版社, 2007.
- [6] 刘文怡, 袁伟涛, 张斌珍. 基于 FPGA 和 USB 的 64 路数据采集系统设计[J]. 微电子学与计算机, 2008, 25(12): 151-153.
- [7] 薛小刚, 葛毅敏. Xilinx ISE 9.x FPGA、CPLD 设计指南[M]. 北京: 人民邮电出版社, 2007.

作者简介:

陈萍女, (1983-), 硕士研究生. 研究方向为飞行器综合电子系统.

姜秀杰女, (1965-), 博士, 研究员. 研究方向为飞行器综合电子系统、飞船、卫星等空间飞行器的研制.

(上接第 124 页)

系统的特征提取、训练算法和模式识别, 大大提高了系统的处理速度. 同时针对语音识别系统的预处理和特征提取阶段, 引入小波变换对传统的实现方法进行改进, 在模式识别中采用了人工神经网络, 有效地降低了系统中的噪声. 减少了系统中不必要的数据量, 有效提高了 DSP 与 MCU 的运行效率和整个语音识别系统的性能.

参考文献:

- [1] 徐秀平, 李柱峰. 基于神经网络的语音识别与自动检验研究[J]. 微电子学与计算机, 2004, 21(12): 123-126.
- [2] 常丹华, 郑春蕾. 基于 DSP 的语音识别智能控制系统[J]. 电子测量技术, 2008, 31(4): 175-178.

- [3] 杜亚玲, 刘建业. 基于 DSP 的航姿系统多传感器信息融合技术[J]. 传感器技术, 2005, 24(9): 85-88.
- [4] 唐述, 刘东, 尹怡欣. 基于粗糙集和神经网络的信息融合方法[J]. 微计算机信息, 2007, 23(6/3): 196-197.
- [5] 李苏梅, 韩国强, 周咏梅. 一种基于 BP 神经网络的颜色空间量化方案[J]. 广西师范大学学报: 自然科学版, 2008, 26(1): 232-235.
- [6] 王岩, 苏健民, 于慧伶, 等. 小波分析在语音识别系统中的研究与应用[J]. 林业劳动安全, 2007, 20(1): 38-41.

作者简介:

宋弘男, (1973-), 硕士, 副教授, 硕士生导师. 研究方向为人工智能、机器人等.