

VC++和 Matlab 混合编程的语音识别研究

Speech Recognition Research based on VC++ and Matlab Mixed Programming

(黄山学院) 华婷婷 宋卫华

HUA Ting-ting SONG Wei-hua

摘要: 采用 VC++ 和 Matlab 混合编程搭建了一个高效的基于 HMM 的语音识别实验验证平台。结合 FPGA 的特点,直接使用加法器、乘法器、比较器等建立一个 Viterbi 算法结构,采用改进方法计算 Viterbi 得分,实现一种简单的基于 HMM 的语音模板匹配。

关键词: 语音识别; Viterbi; VC++; Matlab; FPGA

中图分类号: TN912.34

文献标识码: A

Abstract: Using VC++ and Matlab mixed programming, an efficient HMM-based speech recognition system experiment platform is implemented on an FPGA. Considering FPGA structure characteristic, a Viterbi algorithm structure is realized directly with adder, multiplier, comparator etc. Using modified Viterbi scoring procedure, a simple speech template matching is realized based on HMM.

Key words: Speech Recognition; Viterbi; VC++; Matlab; FPGA

技术创新

1 引言

Matlab 是一款高性能的数值计算和可视化软件,集成数值分析、矩阵计算、信号运算、信号处理和图形显示于一体,构成了一个方便的、界面友好的用户环境。目前,基于 Matlab 的语音识别开发平台虽然在可读性、可移植性和可扩充性上优于其它编程语言,且调试功能强大、数据库函数丰富,可使研究人员“站在巨人的肩上”更加直观、方便地进行分析、计算与设计工作,从而大大地节省了时间。但考虑到其执行代码速度低下,不能直接与硬件底层直接接触等缺点,因此提出了采用 Matlab 和 VC++ 混合编程来搭建语音识别实验平台,并对传统 Viterbi 算法进行变形,直接使用 FPGA 的加法器、比较器和逻辑操作来计算观察值序列,以实现一种简单的嵌入式语音模板匹配。

2 基于 HMM 的语音识别

2.1 语音识别系统

语音识别系统(Speech Recognition System, SRS)基本上是一个模式分类的任务,即通过训练,系统能够把输入的语音按一定模式进行分类。实验在 Matlab 7.0 系统上建立了一个简单的基于隐马尔可夫模型(Hidden Markov Model, HMM)的语音识别过程,如图 1。

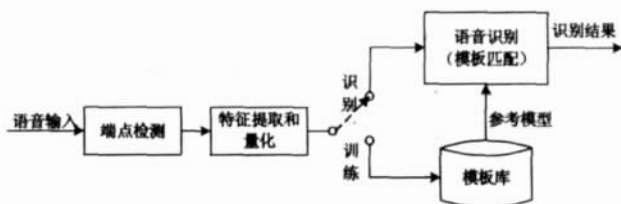


图 1 基于 HMM 的语音识别系统

(1)语音输入:在一般实验室环境下进行语音信号采样,采样格式为 PCM,采样频率 16 KHZ,A/D 的量化精度 8 Bit。然后经过

去噪、预加重、分帧、加窗等处理过程,去掉语音信号中包含的大量冗余信息,加强语音信号的高频共振峰,便于进行频谱分析。

(2)端点检测:考虑到语音信号的录制是在较为安静的实验室环境下进行,利用过零率 Z 来检测清音,用短时能量 E 来检测浊音,两者配合实现可靠的端点检测。

(3)特征提取和量化:对有效语音段进行特征提取,即提取基于 Mel 刻度的倒频谱矢量(Mel Frequency Cepstrum Coefficients, MFCC),它是识别过程中的输入特征值。特征值经矢量量化 Vector Quantization, VQ, 输出 VQ 码本类别号,即 HMM 训练与识别阶段使用的观察值序列 o 。

(4)模型训练与语音识别:训练阶段,系统采用一系列训练观察值估计 HMM 参数,为词汇表中的每个单词建立一个 HMM 模型 $\lambda_v = (\pi_v, A_v, B_v)$, $v = 1, 2, \dots, W$, 包括一个状态转移矩阵 A, 一个观察值输出概率矩阵 B 和一个初始状态概率 π , W 是单词模型的数目。识别阶段,对于给定的观察值序列 o 和 HMM 模型 λ , 识别器采用 Viterbi 算法计算每个 HMM 模型 λ_v 产生 O 的概率 $P(o|\lambda_v)$, 并使得该概率达到最大值,那么该模型所对应的词条标号 v 即为语音识别的结果。

2.2 Viterbi 算法

由于计算复杂度的限制,对于基于 HMM 的实时语音识别来说,需要设计一个高效的硬件结构来执行 Viterbi 译码过程,以加速 HMM 的识别过程。考虑了 FPGA 的特点,分别采用对数概率和状态概率的最小路径对传统的 Viterbi 算法进行变形,其计算 $P(o|\lambda_v)$ 的过程如下:

1. 初始化:当 $t = 1, 1 \leq j \leq N$ 时

$$\delta_1(j) = \pi_j + \log b_j(o_1) \quad (1)$$

2. 递归:当 $2 \leq t \leq T, 1 \leq j \leq N$ 时

$$\delta_t(j) = \min_{i=1, \dots, N} [\delta_{t-1}(i) + a_{ij}] + \log b_j(o_t) \quad (2)$$

3. 结束:当 $t = T$ 时

$$P(o|\lambda_v) = \min_{1 \leq i \leq N} [\delta_T(i)] \quad (3)$$

式中, $a = \{a_{ij}\}$ 为状态 i 转移到状态 j 的概率, N 是状态数目, T 是观察值序列 $o = [o_1, o_2, \dots, o_T]$ 的帧数。

华婷婷: 硕士

通过上面的变形,不仅可以使传统 Viterbi 算法中的乘法转成加法,降低时间消耗,有效地避免数据下溢的问题。而且随着 Viterbi 计算过程的进行,已计算的状态概率值随之增加,改原来找结束概率的最大值为最小值。因此,只需要计算 T 时刻的概率(i),它是大于前参考单词模型的最小值 P_v 的。

实验将直接使用 FPGA 的加法器、比较器和逻辑操作来实现上述公式(2)和公式(3),可以显著提高系统效率,系统结构如图 2。

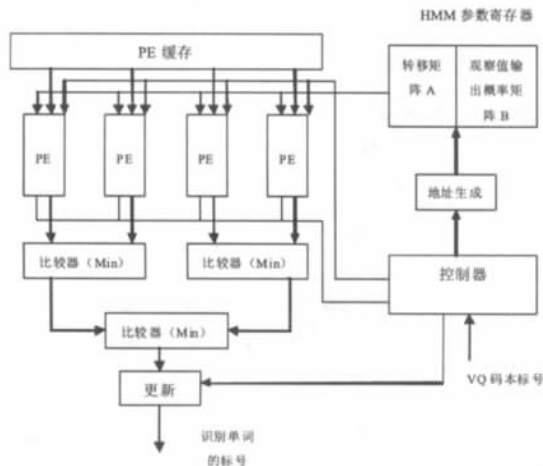


图 2 改进 Viterbi 算法的 VLSI 结构图

在这一方案中,识别过程直接由 FPGA 芯片内的逻辑块从观察序列中计算概率得分,其中,观察值序列通过 VQ 得出。系统包括了两个用来存储转移矩阵 A 和输出概率矩阵 B 的存储器,一个处理单元(Processing Element, PE)阵列,控制器,地址生成和附加比较逻辑。PE 包括有 Viterbi 算法的核心模块—比—选单元(Add-Compare-Select Unit, ACSU),状态累加器,和用来比较 $\delta_t(i)$ 和极值 P_v 的附加比较器。PE 从 HMM 参数寄存器中取出参考模型,沿最小路径计算其概率,然后与极值 P_v 进行比较。当 $\delta_t(i)$ 大于 P_v 时,控制器在下一状态时使 PE 操作无效;同时,控制器控制存储器缓冲操作,并生成整个计算过程中的控制信号。

3 VC++和 Matlab 混合编程

对于在 FPGA 上实现语音识别的核心模块——Viterbi 算法时,有许多工作需要在实验前完成,如定制硬件源代码、转换浮点数据为定点数据和电路仿真等。为减少这部分工作,采用软硬件协同设计的思想,由软件来执行 HMM 模型训练和其它识别过程(如 MFCC、VQ 等)。在实验时,用软件来执行 HMM 模型训练和语音单词识别。然后,把实验数据(语音数据和 HMM 模型参数)转换成定点数据格式,由 PCI 设备驱动程序将实验数据、源代码等下载到硬件,用于 FPGA 验证平台。

根据上述思想,采用 Matlab 和 VC++混合编制 PCI 设备驱动程序,利用 Matlab 系统提供的外部程序调用接口 MEX 文件来实现其于 VC++的混合编程。MEX 文件是一种约定格式编写的文件,使用 C 语言或 FORTRAN 语言编写,是由 Matlab 解释器自动调用并执行的动态链接函数 (Dynamic Link Library Function),它在 Mac 下以 .mex 为后缀名,在 Windows 下即 .dll 文件。基于 C 语言的 MEX 文件主要由两部分组成,第一部分称为入口子程序,其作用是在 Matlab 系统与被调用的外部子程序间建立通信联系。第二部分称为计算功能子程序,它包含所有实际需要的功能的源代码,由入口子程序调用。

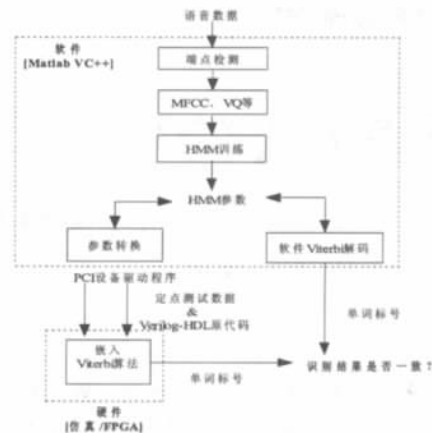


图 3 软硬件协同设计实现嵌入式 Viterbi 算法

该方法可以在软硬件之间达到一致的识别结果,其方案描述如图 3 所示。实验中,计算由 FPGA 硬件完成,该子程序的主要负责 FPGA 与 PCI 的数据传递,即 PCI 设备驱动。通过 MEX 文件,不仅可在 Matlab 系统中像调用内建函数一样调用存在的算法,使资源得到充分利用,避免重复程序设计。同时,还可以对硬件直接进行编程,弥补 Matlab 的不足。

4 实验

该语音识别实验采用的硬件平台是包括有 Altera Cyclone 系列 EP1C12 的 FPGA 和 PCI9054 芯片的 PCI 开发板。EP1C 的 FPGA 负责硬件 Viterbi 计算,PCI9054 在驱动程序的帮助下负责 PC 和 FPGA 间实验数据和结果的传输。

由于 FPGA 的空间限制,实验选择了 4 状态的 HMM 模型和容量 64 的 VQ 码本,占用 FPGA 的 LE(逻辑单元)1,125 个,存储单元占用约 132K 位。然后将 .sof 目标文件下载到 PCI 卡上的 FPGA 芯片中运行,在 Matlab 中调用 VC++编写的 PCI 设备驱动程序,将 VQ 后的语音数据和 HMM 模型参数传送给 FPGA 内的 Viterbi 译码电路,实验中,通过驱动程序输出模板标号与实际语音的标号及仿真实验导出的标号一致。

在 P4 3.0GHz 的 PC 机和 200MHz FPGA 验证平台上,对于约 100 帧的单个语音文件识别而言,软/硬件 Viterbi 算法的耗时如下表 1 所示。

表 1 软/硬件实现 Viterbi 算法耗时比较

	Matlab (3.0GHz)	FPGA (200MHz)
耗时 (ms)	14.4	8

由上述实验结果证明了该 Viterbi 算法的 VLSI 结构能够准确且快速地完成语音识别的解码过程,满足嵌入式计算精度要求,表明该实现方案是切实可行的。

5 结束语

本文的创新点:采用 Matlab、VC++和 FPGA 搭建了一个软硬件协同的语音识别实验研究平台,以 VC++来弥补 Matlab 不能与硬件底层进行直接接触的不足。并在传统 Viterbi 算法基础上,对其采取一定变形,直接使用 FPGA 的加法器、比较器和逻辑操作建立 Viterbi 算法的 VLSI 结构,来计算观察值序列,以实现一种简单的基于 HMM 语音识别的模板匹配。采用这种软硬件协同的实验研究平台,可在利用前面 Matlab 的实验成果基础上,逐步实现语音识别各功能模块的嵌入式设计,减少工作量,并易于调试。

(下转第 99 页)

- [3]Perelson A. Immune network theory. Immunological Review, 1989, 110:5-36
- [4]Farmer J D, Packard N H, Perelson A S. The immune system, adaptation, and machine learning. Physical D, 1986, 22:187-204
- [5]Ishida Y, Fully Distributed Diagnosis by PDP Learning Algorithm: Towards Immune Network PDP Model. Proceedings of IC-NN90, San Diego, 1990
- [6]Kayawa M, Sugita Y, Morooka, Sensor Diagnosis System Combining Immune Network and Learning Vector Quantization. Electrical Engineering in Japan, 1997:44-55
- [7]Dasgupta D, KrishnaKumar K, Wong D, Berry M, et al. Negative selection algorithm for aircraft fault detection [oL]. <http://issrl.cs.memphis.edu/papers/ais/2004/ICARIS04.pdf>
- [8]Dasgupta D, Forrest S. Artificial immune systems in industrial applications [C]//Proc. 2nd International Conference on Intelligent Processing and Manufacturing of Materials, Honolulu, 1999:257-267.
- [9]Gonzalez F, Dasgupta D. Anomaly detection using real-valued negative selection [J]. Genetic Programming and Evolvable Machines, 2003, 12(4): 383-403.
- [10]Ishiguro A, Watanabe Y, Uchikawa Y. Fault diagnosis of plant systems using immune networks. In: Proc IEEE International Conference On multi-sensor Fusion and Integration for Intelligent Systems, Las Vegas, NV, 1994, 34-42
- [11]Taniguchi S, Dote Y. Sensor fault detection for uninterruptible power supply control systems using fast fuzzy neural network and immune network[C]//Proc. of the SMC'2001, USA, Oct, 2001: 7-10.
- [12]Tang Z, Yamaguchi T, Tashima K et al. Multiple-valued immune network model and its simulation. In: Proc the 27 International Symposium on Multiple-Valued Logic, Antigonish, Nova Scotia, Canada, 1997, 519-524
- [13]Luh Guan-chun, Cheng Weichong. Immune model-based fault diagnosis[J]. Mathematics and Computers in Simulation, 2004, 7: 5-6.
- [14]刘树林, 黄文虎, 王日新等. 基于免疫系统的往复泵在线故障诊断方法. 中国机械工程学报, 2002, 13(8): 686-689
- [15]谷吉海, 姜兴渭, 刘树林等. 免疫系统的反面选择算法在故障诊断中的应用. 中国空间科学技术, 2002(2): 24-29
- [16]栾家辉, 姜兴渭. 免疫状态观测器的设计及应用. 中国空间科学技术, 2005, 5: 34-39
- [17]庞茂, 周晓军, 孟庆华. 基于免疫学的在线故障检测算法的研究及应用. 中国电机工程学报, 2005, 25(24): 149-153
- [18]李伟, 黄席樾. 基于免疫原理的故障诊断推理模型研究. 计算机仿真, 2005, 22(7): 111-113
- [19]陈强, 郑德玲. 基于免疫原理的齿轮箱故障检测和诊断方法研究. 矿山机械, 2005, 33(5): 75-77
- [20]李蓓智, 杨建国, 杨江云, 等. 基于自我-非我识别机理的状态监测与故障诊断[J]. 上海工程技术大学学报, 2004, 18(1): 24-27.
- [21]窦唯, 于楷, 孟庆武, 等. 基于免疫系统机理的距离函数故障诊断方法[J]. 大庆石油学院学报, 2003, 27(3): 54-56.
- [22]窦唯, 刘树林, 孙明, 等. 生物免疫机理在往复压缩机在线状态监测中的应用[J]. 流体机械, 2004, 32(5): 16-19.
- [23]殷桂梁, 肖丽萍, 吴长奇, 等. 免疫原理用于异步电动机故障诊断的研究[J]. 中国电机工程学报, 2003, 23(6): 132-136.
- [24]孟庆华, 周晓军, 吴跃成, 等. 基于小波免疫系统的车辆总成故障检测[J]. 汽车工程, 2004, 26(5): 619-622.
- [25]唐志航, 何宏, 胡志望, 等. 改进的 BP 神经网络在故障诊断中的应用[J]. 微计算机信息, 2008, 5-1: 171-173.

作者简介: 邵龙秋(1984-), 男(汉族), 浙江温州人, 太原理工大学在读研究生, 研究领域为人工免疫系统, 故障诊断。

Biography: SHAO Long-qiu (1984-), male (the Han nationality), Wenzhou city Zhejiang province, postgraduate of Taiyuan University of Technology, the researcher direction: artificial immune system, fault diagnosis.

(030024 山西太原 太原理工大学信息工程学院) 邵龙秋 李铁鹰 (525000 广东茂名 茂名学院计算机与电子信息学院) 邵龙秋 张清华 张亚社

(030024 山西太原 太原理工大学计算机与软件学院) 张亚社 (College of Information Engineering, Taiyuan University of Technology, Taiyuan Shanxi 030024, China) SHAO Long-qiu LI Tie-ying

(College of Electronic Information and Computer, Maoming University Maoming, Guangdong, 525000, China) SHAO Long-qiu ZHANG Qing-hua ZHANG Ya-she (College of Computer and Software Engineering, Taiyuan University of Technology, Taiyuan Shanxi 030024, China) ZHANG Ya-she

通讯地址: (525000 广东茂名市 茂名学院(官渡校区) 硕士研究生信箱) 邵龙秋

(收稿日期: 2008.11.24) (修稿日期: 2009.02.24)

(上接第 175 页)

参考文献

- [1] 丁元力. Matlab 语言在数字语音处理上的应用[J]. 电声技术, 2001, 9: 7-9
- [2] 刘加, 刘润生. 语音识别技术及应用[J]. 世界电子元器件, 2001, 11: 21-23
- [3] 覃溪, 郑建华等. 一种语音端点检测的方法及改进[J]. 微计算机信息, 2007, 8-3: 231-233
- [4] Shingo Yoshizawa, Naoya Wada, Noboru Hayasaka. VLSI architecture for HMM-based speech recognition systems and its verlatplatform [C]. International Symposium on Communications and Information Technologies 2004, 700-703
- [5] 韩纪庆, 张磊, 郑铁然. 语音信号处理[M]. 清华大学出版社, 2004, 200-218
- [6] Bok-Gue Park, Koon-shik Cho, Jun-Dong Cho. Low power VLSI architecture of Viterbi scorer for HMM-based isolated word recognition [C]. Proceedings International Symposium on Quality Electronic Design, 2002, 235-239

作者简介: 华婷婷(1981-), 女, 汉, 江苏镇江, 黄山学院信息工程学院, 硕士, 主要研究: 语音识别、嵌入式应用; 宋卫华(1983-), 女, 汉, 河南商丘, 黄山学院信息工程学院, 硕士, 主要研究: 医学图像检索。

Biography: HUA Ting-ting (1981-), female (han ethnic), jiangsu province, Department of Computer Science & Engineering Huangshan University, master, research in Speech Recognition, Embedded Application.

(245021 安徽黄山 黄山学院信息工程学院) 华婷婷 宋卫华 (Department of Computer Science & Engineering Huangshan University, Huangshan, 245041, China) HUA Ting-ting SONG Wei-hua

通讯地址: (245021 安徽省黄山市屯溪区戴震路 44 号信息工程学院) 华婷婷

(收稿日期: 2008.11.20) (修稿日期: 2009.02.20)