

基于 MATLAB 和 BP 网络的语音识别系统

The speech recognition system based on MATLAB and BP neural networks

(华南理工大学) 詹新明 杨 灿

ZHAN Xin-ming YANG Can

摘要: 神经网络是近年来信息科学、脑科学、神经心理学等诸多学科共同关注和研究的热点。由于其具有良好的抽象分类特性,现已应用于语音识别系统的研究和开发,并成为解决识别相关问题的有效工具。文章在讲述语音识别过程的基础上重点讨论利用 BP 神经网络对语音进行识别,用 MATLAB 完成对神经网络的训练和测试,并获得满意的结果。

关键词: MATLAB; BP 神经网络; 语音识别

中图分类号: TN912.34 **文献标识码:** A

Abstract: In recent years neural network is study hotspot of common attention on subject of information science, brain science, neural psychology and so on. Because neural network has a good abstract classification characteristic, now it has been used in the research and development of speech recognition system, and becomes the valid tool which resolves a related problem gradually in speech recognition. This paper briefly introduces process of speech recognition and emphatically discusses how to recognize speech with BP Neural Networks, trains and tests the neural network in use of MATLAB, and acquires a satisfied result.

Key words: MATLAB; BP neural networks; Speech recognition

1 引言

人工神经网络(Artificial Neural Network)是一种模拟人脑神经元细胞的网络结构和功能,运用大量的处理部件,由人工方式建立起来的自适应非线性动态系统。由于它模拟了人类神经元活动的原理,具有自学习、联想、对比、分类、推理和概括能力,因此被应用于信息处理和语音识别等领域。数字化的语音信号可作为一维或二维(双声道立体声数据)矩阵来处理,而 MATLAB 的最大特点和优势就是矩阵运算的能力很强,因此很自然地,将 MATLAB 应用到语音处理领域。基于这种前提和背景,将 MATLAB 和 BP 神经网络相结合进行语音识别的研究,已成为语音识别研究领域的又一个热点。本文主要介绍将语音信号经过识别系统预处理后进行特征提取,再利用 BP 神经网络进行训练和测试,最后得出识别结果,并给出了基于 MATLAB 环境下的图形用户界面 GUI 的实现。

2 BP 神经网络基本原理

基于误差反向传播(Back Propagation)算法的多层前馈网络(Multiple-layer feed forward network),简称为 BP 网络,又称误差信号反馈网络,是神经网络的一个分支,也是目前神经网络中应用最多也是最成功的网络之一。它是一种有教师的学习网络,能够实现从 N 维到 M 维的非线性映射,一般采用梯度下降法实现快速收敛。而构造一个 BP 网络需要确定其处理单元,即神经元的特性和网络的拓扑结构。

2.1 BP 网络的拓扑结构

神经网络的拓扑结构是指神经元之间的互连结构。BP 神经网络采用的是并行网络结构,包括输入层、隐含层和输出层,经作用函数后,再把隐节点的输出信号传递到输出节点,最后给

出输出结果。由图 2 可见各层次的神经元之间形成全互连接,各层次内的神经元之间没有连接。这种结构使多层前馈网络可在输入和输出间建立合适的线性或非线性关系,又不致使网络输出限制在-1 和 1 之间。图 1 是一个三层的 BP 网络结构。

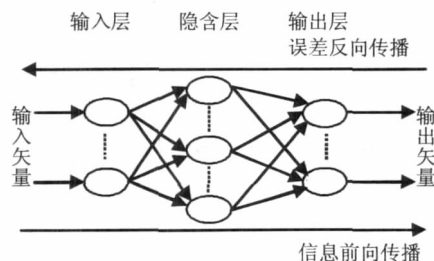


图 1 BP 网络结构

2.2 反向传播算法

算法分为两个阶段:

第一阶段(正向过程)输入信息,从输入层经隐层逐层计算各单元的输出值;根据下式计算每层的输出值:

$$net_j = \sum w_{ij} O_i \quad (1)$$

$$O_j = f(net_j) \quad (2)$$

第二阶段(反向传播过程)输出误差,逐层向前算出隐层各个单元的误差,并用此误差修正前层的值。在 BP 算法中常采用梯度法修正权值,为此要求输出函数可微,通常采用 Sigmoid 函数作为输出函数。

$$(1) \text{ 计算误差: } E = \frac{1}{2} \sum_j (y_j - \hat{y}_j)^2 \quad (3)$$

(2) 按照梯度方向计算各层权重的修正量:

$$\Delta w_{ij}(t) = \frac{\partial E}{\partial w_{ij}} = -\eta \delta_j O_i \quad (4)$$

其中 δ_j 的计算公式为:

$$\delta_j = \begin{cases} -(y - O_j) O_j (1 - O_j) & \text{输出层的计算方法} \\ O_j (1 - O_j) \sum_k w_{jk} \delta_k & \text{非输出层的计算方法} \end{cases} \quad (5)$$

詹新明: 硕士研究生

(3)修正各层权重:

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t) \quad (6)$$

(4)重复上述过程直到算法收敛,即误差小于给定的阈值。

3 语音识别系统原理

语音识别系统的典型识别方案如图2所示。这是一个基于模式匹配原理的自动语音识别系统方框图,也是目前大多数语音识别系统都采用的设计原理。根据这个原理,未知语音的模式要与已知语音的参考模式逐一进行比较,选出最佳匹配的参考模式作为识别结果。

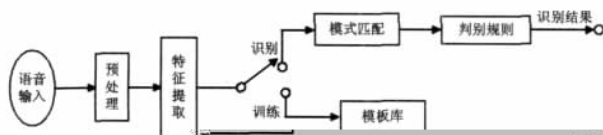


图2 语音识别系统原理图

图中,待识别语音先经话筒转换成语音信号,或直接从文件中读取,然后从识别系统前端输入,再进行预处理。这时人的语音信号还只是模拟信号,必须经过预处理,转化为数字信号。预处理包括预滤波、采样和量化、分帧加窗、端点加重等,有时还包括模数转换器。特征提取部分用于语音中反映本质特征的声学参数,常用的特征有短时平均幅度、短时平均过零率、短时自相关函数、线性预测系数/浊音标志、基音频率、短时傅立叶变换、倒谱、共振峰。在识别之前进行,是通过多次重复语音的方式从原始语音中去除冗余信息,保留关键数据,再按一定规则对数据加形成模式库。模式匹配是整个语音识别系统的核心,是选定的准则以及专家知识(如构词规则、语法规则、语义规则)输入特征与库存模式之间的相似度,选择最大相似度识别结果,从而判断出输入语音的语意信息。

4 MATLAB 环境下语音识别系统的

分帧加窗

语音信号是一种典型的非平稳信号,随时间变化而变化,但音的形成过程是与发音器官的运动密切相关的,因此在一定的时间内,语音信号的频谱特性和物理特征参量可作是不变的,即可视为短时平稳的,就可以对语音信号进行处理。

帧相当于对原始信号加了一个矩形窗。用 enframe 进行相当于默认加一个矩形窗。加窗的目的就是使信号在重新时候,频谱特性和原来比较接近。对语音信号的时域分析函数的形状是非常重要的,矩形窗的谱平滑性较好,但波节丢失,并且矩形窗的旁瓣较大会产生泄漏现象,而汉明有效的克服这一点,因此,用 enframe 分帧后,最好加一个 g(汉明)窗,分析效果会更好。

端点检测

语音信号的端点检测可以分为四段:静音、过渡段、语音段、静音。文用典型的双门限算法,在程序设计中以变量 status 来前所处的状态。在静音段,如果能量或过零率超过了低门限该开始标记起始点,进入过渡段。在过渡段中,只要两个数值都回落到低门限以下,就将当前状态恢复到静音状态。如果在过渡段中的两个参数中的任意一个超过了高门限,

就可以确信进入语音段了。以下是该设计的关键代码:

I 当 status 为 0/1 时:

(1)满足 $\text{amp}(n) \geq (1.3 * EN) \& (\text{zcr}(n) \geq (0.3 * ZR) \& \text{zcr}(n) \leq 4.5 * ZR)$ 时,进入状态 2;

(2)满足 $\text{amp}(n) \geq (0.10 * EN) \& (\text{zcr}(n) \geq (0.10 * ZR))$ 时,进入状态 1;

(3)否则进入状态 0。

II 当 status 为 2 时:当不满足 $\text{amp}(n) \geq (0.007 * EN) \& (\text{zcr}(n) \geq (0.007 * ZR))$ 时,则进入状态 3。

考虑到一些突发性的噪声也可以引起短时能量或过零率的数值很高,但由于其往往不能长时间维持,所以,可以通过设定最短时间门限来判别。当前状态处于语音段时,如果两个参数的数值降低到低门限以下,且总的计时长度小于最短时间门限,则认为这是一段噪声,继续扫描以后的语音数据。否则就标记好结束端点,并返回。

4.3 特征参数的提取

语音信号是一种典型的时变信号,特征提取的目的就是对原始语音进行处理后,计算语音对应的特征参数。提取的特征参数满足如下要求:

(1)特征参数能有效地代表语音特征,具有很好的区分性。

(2)参数间有良好的独立性。

(3)特征参数要计算方便,要考虑到语音识别的实时实现。

语音特征参数可以是能量、基音频率、共振峰值等语音参数,目前在语音识别中较为常用的特征参数为线性预测倒谱系数(LPCC)与 Mel 倒谱系数(MFCC)。本文采用 MFCC 特征参数,其提取及计算步骤如下:

(1)原始语音信号 $s(n)$ 经过预加重、分帧、加窗等处理,得到每个语音帧的时域信号 $x(n)$ 。

(2)将时域信号 $x(n)$ 后补若干 0 以形成长为 N (一般取 $N=512$) 的序列,然后经过离散傅立叶变换(DFT)后得到线性频谱 $X(k)$,转换公式为

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi nk/N} \quad (0 \leq n, k \leq N-1) \quad (7)$$

在实际应用中,常常通过快速傅立叶变换(FFT)过程加以计算,其中 N 一般称之为 DFT(或 FFT)窗宽。

(3)将上述线性频谱 $X(k)$ 通过 Mel 频率滤波器组得到 Mel 频谱,并通过对数能量的处理,得到对数频谱 $S(m)$ 。

(4)将上述对数频谱 $S(m)$ 经过离散余弦变换(DCT)变换到倒频谱域,即可得到 Mel 频率倒谱系数(MFCC 参数) $c(n)$:

$$c(n) = \sum_{m=1}^{M-1} S(m) \cos\left[\frac{\pi n(m+1/2)}{M}\right] \quad (0 \leq n < M) \quad (8)$$

5 仿真实验

本实验的样本库为 200 个 wav 文件(含男女各 50 人,每人两次 0~9 的录音),音频格式为音频格式为 PCM、22kHz、16 位、双声道,训练网络采用 3 层 BP 网络。实验过程中,先对其中 100 个样本进行训练,然后再用其它的样本进行识别,随着隐层神经元数目的改变,步数和识别率也随着改变,从中选择最适当的神经元数,作为训练网络的神经元数目。本实验中,还加入动量因子,对 BP 算法进行改进,使其收敛更快。利用本系统对样本库进行训练和识别,平均识别率为 97.8%,达到了预期的目标。图 3 为识别结果界面。

通过实验,当隐层神经元个数选定为 25 时,BP 网络在训练时间、实时反映速度及识别率上都有明显的提高。表 1 为改进后 BP 网络识别结果。

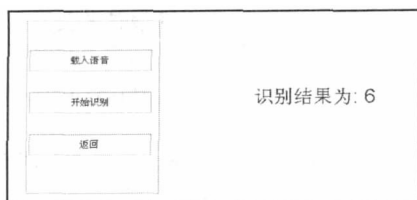


图3 识别结果

表1 改进后 BP 网络识别结果

隐层神经元个数	30	25	20	15
训练步数	476	326	638	521
识别率	99%	100%	97%	95%

6 结论

本文作者的创新点在于充分利用 MATLAB 矩阵运算能力强及可视化特性,并结合 BP 神经网络,设计了针对 0~9 等数字的语音识别系统,使语音识别的每个过程实现可视化,达到了期望的识别率,而且还引入动量因子对 BP 网络加以改进,进一步提高了识别的实时性和识别率。

参考文献

- [1]杨健刚.神经网络实用教程[M].杭州:浙江大学出版社,2001,01.
- [2]王家文,王皓,刘海.MATLAB7.0 编程基础[M].北京:机械工业出版社,2005,07.
- [3]Duane Hanselman, Bruce Littlefield. Mastering MATLAB: a comprehensive tutorial and reference. Prentice[M]Hall,1996.
- [4]蒋宗礼.神经网络导论[M].北京:高等教育出版社, 2001.
- [5]王炳燭,屈丹,彭煊等.实用语音识别基础[M]. 北京:国防工业出版社,2005,01.
- [6]易克初,田斌,付强.语音信号处理[M].北京:国防工业出版社, 2001:150~160.
- [7]江官星,王建英.一种改进的检测语音端点的方法[J].微计算机信息,2006,5-1:138~139.
- [8]蔡莲红,黄德智,蔡锐.现代语音技术基础与应用[M].北京:清华大学出版社,2003.

作者简介:詹新明(1978-),男,汉族,湖北浠水人,华南理工大学计算机科学与工程学院计算机应用技术专业硕士研究生,主要研究方向:多媒体通信、语音识别; 杨灿(1971-),男,汉族,湖南沅江人,华南理工大学计算机科学与工程学院副教授,博士,硕士研究生导师,主要研究方向:多媒体通信,新媒体技术,P2P 流媒体,分布式计算。

Biography:ZHAN Xin-ming (1978-),male, Hubei, School of Computer Science and Engineering, South China University of Technology, Post-graduate, Major in Computer Application Technology, Engaged in the study of multi-media communication and speech recognition.

(510006 广州 华南理工大学计算机科学与工程学院) 詹新明 杨 灿
(School of Computer Science and Engineering, South China University of Technology, Guangzhou, 510006, China)
ZHAN Xin-ming YANG Can

通讯地址:(510006 广州 广东省广州市番禺区大学城中环东路华南理工大学 C14 栋 224 室) 詹新明

(收稿日期:2008.12.01)(修稿日期:2009.03.01)

(上接第 156 页)

- [2]陈根社,朱建民,陆彦.飞机悬挂物管理系统的仿真研究[J].西北工业大学学报,1993,11(2):199~200

- [3]陆彦,陈根社.飞机外挂物管理系统研究[J].电光与控制,1991,(1):1~4
- [4]田宏理,黄长强,唐玉棋,罗新成.空对地武器选用方法[J].火力与指挥控制,2006,31(3):58~60.

- [5]周巍,马振江,夏国平,颜宁.导弹武器效能评估系统的设计与实现[J].微计算机信息,2007,11-3:33~34

作者简介:韩广奇(1979-),男,汉,河南洛阳人,硕士研究生,研究方向为航空综合火力控制原理及控制与仿真。

Biography:HAN Guang-qi (1979-),male (han),born in Luoyang of Henan, a postgraduate in Air Force Engineering University, Research area: fire-control theory and simulation.

(710038 西安 空军工程大学工程学院) 韩广奇 于 雷 周中良
(Engineering Institute,Air Force Engineering University, Xi'an, 710038,China) HAN Guang-qi YU Lei ZHOU Zhong-liang
通讯地址:(710038 陕西西安 空军工程大学工程学院二系火控教研室) 韩广奇

(收稿日期:2008.12.19)(修稿日期:2009.03.19)

(上接第 173 页)

从图 2、图 3 中可以看出,采用惯组、测速仪组合导航方式,通过卡尔曼滤波仿真计算后,速度、位置误差非常小,路线变化图与实际跑车路线几乎一致,出发路线与返回路线在同一路面时基本重合,误差在 20m 以内,起点与终点的位置之差也在 20m 以内,达到较高的导航精度。

5 结论

本文作者创新点:建立了组合导航系统卡尔曼滤波误差模型。对比纯惯组导航计算和组合导航计算,可以看出,采用 IMU、测速仪组合导航方式,通过闭环卡尔曼滤波计算,定位精度在 20m 以内,验证了本文所建模型的正确性。

参考文献

- [1]董绪荣,张守信,华仲春.GPS/INS 组合导航定位及其应用.长沙:国防科技大学出版社,1998.
- [2]袁信,俞济祥,陈哲.导航系统.北京:航空工业出版社,1992
- [3]刘基干,李征航,王跃虎等.全球定位系统原理及应用.北京:测绘出版社,1993
- [4]李涛,吴美平,江明等.激光陀螺 SINS/GPS 组合导航车载试验系统.国防科技大学学报,2003,25(3):104~108
- [5]傅惠,徐建闽,基于卡尔曼滤波的路径行程时间预测方法.微计算机信息,2007,8-1:290~292
- [6]刘基余.GPS 卫星导航定位原理与方法.北京:科学技术出版社,2003
- [7]张其善,吴今培,杨东凯.智能车辆定位导航系统及应用.北京:科学技术出版社,2002
- [8]郑贵省.GPS/DR 车载组合定位系统数据融合算法研究.[博]天津:天津大学,2005

作者简介:王志坤,男,1978 年出生,工程硕士,主要从事惯性器件、惯性导航等领域的研究。

Biography:WANG Zhi-kun, 1978, the Han nationality, Work in China Aerospace Science and Industry Corporation7801,study in National University of Defense Technology, engineering master, major in: Inertial device, inertial navigation etc.

(410205 中国航天科工集团 7801 研究所) 王志坤 李永满 蔡铁鑫

(410073 国防科学技术大学) 王志坤

通讯地址:(410205 湖南省长沙市 585 信箱 3 分箱) 王志坤

(收稿日期:2008.11.26)(修稿日期:2009.02.26)