

小波分析和粒子群优化神经网络的语音端点检测

陈雪芳

(东莞理工学院 计算机学院, 广东 东莞 523808)

摘 要: 为了提高语音端点检测的适应性和鲁棒性, 提出一种小波分析和粒子群优化神经网络(WA-PSO-BP)的语音端点检测算法. 首先利用小波分析提取语音信号的特征量, 然后将特征量作为输入 BP 神经网络进行学习, 并采用粒子群算法优化 BP 神经网络参数, 从而建立语音端检测模型. 仿真结果表明, WA-PSO-BP 提高了语音端点检测正确率, 有效降低了虚检率和漏检率. 这说明 WA-PSO-BP 是一种可行性较高, 环境适应性较强的语音检测算法.

关键词: 小波分析; 神经网络; 语音端点; 粒子群优化算法; 特征选择

中图分类号: TN91

文献标识码: A

文章编号: 1000-7180(2013)09-0094-04

Wavelet Analysis and the Neural Network of Particle Swarm Optimization for Speech Endpoints Detection

CHEN Xue-fang

(School of Computer, Dongguan University of Technology, Dongguan 523808, China)

Abstract: In order to improve the adaptability and robustness of speech endpoint detection, this paper proposes a speech endpoint detection method base on WA-PSO-BP. Firstly, the feature of speech signals are extracted by wavelet analysis; then the features are input to BP neural network to build the speech endpoints detection model in which the BP neural network's parameters are optimized by particle swarm optimization algorithm. The experiments results show that the proposed method has improved the detection rate, and it has better adaptability and robustness. WA-PSO-BP has high reliability makes it suitable for application in different kinds of environments.

Key words: wavelet analysis; neural network; speech endpoints; particle swarm optimization algorithm; feature selection

1 引言

语音端点检测的目的就是在复杂应用环境下, 从信号流中分辨出语音信号和非语音信号, 并确定语音信号的开始及结束. 当前语音检测算法主要有时域和频域两类^[1]. 时域检测算法主要有短时能量、短时平均过零率和短时相关分析等^[2-3], 该类方法在低信噪比环境下失去了作用, 不能进行正确的端点检测, 然而在实际语音检测的应用中, 很难保证有足够高的信噪比, 这样, 时域方法容易将噪声误当成语音信号. 频域检测算法有倒谱、信息熵、谱熵和频带

方差等, 该类检测方法在理想条件下可以获得较好的检测效果^[4-5], 但在恶劣环境下, 该类端点检测准确性仍然很差. 为了获取更优的语音端点检测效果, 学者们提出了特征组合的端点检测算法, 显著提高在低信噪比、噪声动态化的各种环境下的端点检测性能^[6]. 传统语音端点检测器基本上均是基于线性建模方法, 对环境噪声变化不够鲁棒, 虚检率和漏检率比较高.

为了提高语音端点检测正确率, 本文提出一种小波分析(wavelet analysis, WA)、粒子群优化算法(particle swarm optimization algorithm, PSO)和

收稿日期: 2012-11-08; **修回日期:** 2012-12-14

基金项目: 国家自然科学基金(2000927001); 东莞市高校科研机构科技计划项目(2011108102016)

BP神经网络(BP neural network, BPNN)相结合的强噪声背景下语音端点检测方法(WA-PSO-BP).

2 WA-PSO-BP的端点检测系统结构

小波分析(wavelet analysis, WA)在时域和频域都能表征信号的局部特征,对信号的高频部分有较好的时间分辨率,对低频部分有较好的频率分辨率,能够满足语音信号处理的要求.因此可以采用小波分析提取语音信号的特征量. BP神经网络具有很强的自适应、并行性、鲁棒性、非线性逼近能力,因此利用BP神经网络作为语音端点检测器,对小波变换提取到的语音特征量进行训练,同时采用粒子群算法(particle swarm optimization algorithm, PSO)对BP参数进行优化,建立最优语音端点检测模型.基于WA-PSO-BP的语音端点检测系统结构如图1所示.

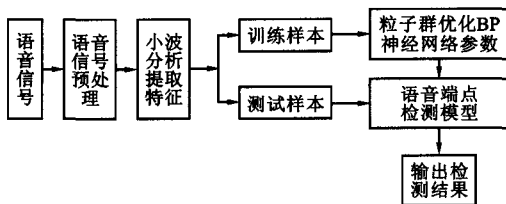


图1 WA-PSO-BP的语音端点检测系统

3 语音端点检测模型

3.1 语音信号的预处理

语音信号是一种非平稳信号,不能直接用于提取特征,对采集语音信号进行预处理,英文单词“eat”的预处理结果见图2.

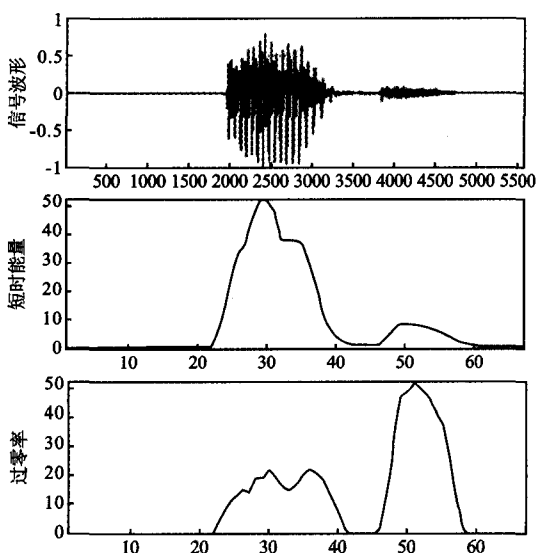


图2 英文单词“eat”的预处理

3.2 语音特征量提取

采用Mallat算法对每帧信号 $f(n)$ 进行5层小波分解,即:

$$\begin{aligned} f(n) &\rightarrow a_1(n) \rightarrow a_2(n) \rightarrow \cdots \rightarrow a_5(n) \\ &\searrow \quad \searrow \quad \searrow \quad \searrow \\ d_1(n) & d_2(n) d_3(n) \cdots d_5(n) \end{aligned} \quad (1)$$

根据小波多分辨率分析原理可知, $d_1(n) \sim d_5(n)$ 、 $a_5(n)$ 可表征原信号所有的频率信号,计算各层小波子带系数的平均能量:

$$E_i = \frac{1}{N} \sum_{n=1}^N |s_i(n)| \quad (2)$$

式中, $s_i(n)$ 代表 $d_1(n) \sim d_5(n)$ 、 $a_5(n)$ 小波子带中的某个子带的小波系数, N 为小波系数的个数^[12].

计算6个小波子带平均能量的均值 E_m 和方差 σ^2 :

$$\begin{cases} E_m = \frac{1}{6} \sum_{i=1}^6 E_i \\ \sigma^2 = \frac{1}{6} \sum_{i=1}^6 (E_i - E_m)^2 \end{cases} \quad (3)$$

共获得8个小波子带平均能量 $E_1 \sim E_6$,一个 E_m 和一个 σ^2 ,它们组成的特征矢量为:

$$X(n) = [E_1, E_2, E_3, E_4, E_5, E_6, E_m, \sigma^2]^T \quad (4)$$

3.3 BP神经网络以及参数优化

BP神经网络是一种包含输入层、隐含层和输出层的多层前馈型网络,大量实践证明三层神经网络能够以任意精度逼近任何的非线性函数,BP神经网络结构如图3所示.

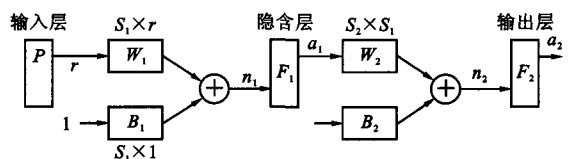


图3 BP神经网络结构

BP神经网络预测性能优劣与其网络连接权值密切相关,传统动量项法或经验法难以获得全局最优连接权值,使网络学习速率小,没有足够的能量跃过局部极小值点,导致语音端点检测正确率低.为了解决BP神经网络存在的难题,采用智能优化搜索方法—粒子群优化算法对其网络连接权值进行优化,获得性能更优的BP神经网络,具体流程如图4所示.

3.4 BP神经网络的语音检测步骤

Step1:初始化BP神经网络、小波分析和PSO相关参数.

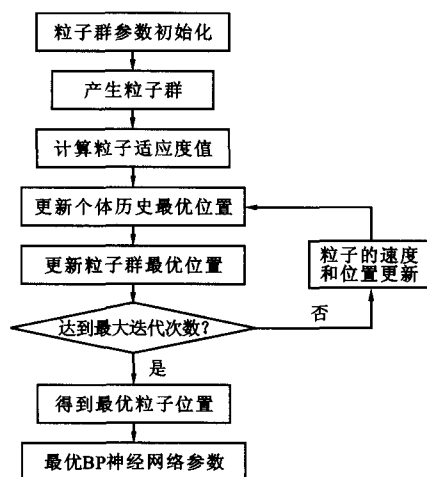


图4 BP神经网络参数优化流程

Step2: 随机选择 1 段含语音段和噪声的信号, 小波分析提取特征量作为 BP 神经网络输入, 人工标示出每帧信号作为 BP 神经网络输出, 完成 BP 神经网络训练样本构建。

Step3: 将训练样本输入 BP 神经网络进行训练, 由于随机确定 BP 神经网络初始参数会导致检测结果与实际结果间差异较大, 因此通过 PSO 对 BP 神经网络优化, 使 BP 神经网络输出与理想输出值达到预设要求, 从而完成 BP 神经网络训练。

Step4: 用训练好的 BP 神经网络对原来训练样本进行计算, 并输出检测结果, 门限值为 0.5, 输出结果大于 0.5, 认为当前帧为语音帧, 否则为非语音帧, 然后将实际输出结果与标示好的信号语音帧进行比较, 如果神经网络训练效果不好, 需要重新对神经网络进行训练。

Step5: 进行语音端点检测。取一段语音信号, 提取其特征量, 然后采用训练好的 BP 神经网络对其进行检测, 最终输出语音端点检测结果。

4 仿真实验

4.1 仿真环境

为了检验 WA-PSO-BP 算法对语单端点检测有效性, 在 CPU P4 3.0GMHZ, 内存 2GB, Windows XP 平台下, 采用 Matlab 2007 软件实现仿真实验。为了使 WA-PSO-BP 算法的检测结果更具有说服力, 选择短时能量特征 + PSO - BP 神经网络 (BPNN1), 小波分析 + BP 神经网络 (BPNN2)、小波分析 + 线性模型 (WA-LINER) 进行对比实验。采用正确率、虚检率、漏检率和检测速度作为算法性能的评价标准。

4.2 数据来源

实验使用带噪语音是由 aurora2.0 语音库中的 130 个干净声音和 Noisex 92 噪声库中的 7 种噪声混合而成 wav 格式的声音, 每个干净语音分别混合 destroyerengine 噪声、pink 噪声、volvo 噪声、white 噪声、bueeaneer2 噪声等 5 种噪声。声音的位速为 128kbPs, 音频采样大小为 16 位, 频道为单声道, 音频的采样级别为 8kHz, 音频格式为 PCM。

4.3 结果与分析

4.3.1 不同语音端点算法的检测性能对比

不同信噪比的条件下, 对比各种语音端点检测算法的仿真结果可以得到:

(1) 对于 white 噪声语音信号进行检测, 各检测算法均获得了比较满意的检测效果, 检测正确率比较高。当噪声环境为 estroyerengine 噪声、pink 噪声、volvo 噪声、bueeaneer2 噪声, 全部语音端点检测算法性能均有所下降, 尤其是 WA-LINER 下降十分明显, 而基于 BP 神经网络的检验正确率仍然较高, 说明 BP 神经网络具有很强的自适应、并行性、鲁棒性, 对语音环境变化适应性比较强。

(2) 相对于 BPNN1, WA-PSO-BP 的检测性能要优。WA-PSO-BP 利用小波分析和 BP 神经网络相结合, 达到了优势互补, 表现出更强的抗噪性, 提高语音端点检测正确率的同时, 有效降低了端点漏检率和虚检率。

(3) 相对于 BPNN2 算法, WA-PSO-BP 检测性能也更优。这说明采用 PSO 对 BP 神经网络的参数进行优化, 可以很好的克服 BP 神经网络因参数选择不当存在的缺陷, 使 BP 神经网络的性能达到全局最优, 进一步提高了音端点检测正确率。

4.3.2 算法检测速度比较

采用运行时间来衡量语音端点检测速度, 在 Matlab2007 平台下, 利用 Tic 和 Toc 命令记录预测模型的训练和预测时间, WA-PSO-BP、BPNN1、BPNN2 和 WA-LINER 的检测时间如表 1 所示。

表1 各种算法的检测速度对比

算法	训练时间(s)	测试时间(s)
WA-LINER	10.55	1.15
BPNN1	9.177	0.66
BPNN2	7.685	0.58
WA-PSO-BP	4.10	0.45

从表 1 可知, 相对于对比算法, 无论是训练时间还是检测时间, WA-PSO-BP 均是最短的, 这说明采用小波分析提取语音特征, 以减少 BP 神经网络的

输入变量数目,降低了计算复杂度,同时采用 PSO 对模型参数进行优化,加快了学习效率,对比结果表明,WA-PSO-BP 更适用于实时性要求比较高的语音端点检测。

5 结束语

端点检测是语音识别中极其重要的一项技术,其准确性对识别性能具有很大影响,在当信噪比较小环境下,传统检测算法难以对语音端点进行正确检测,为此提出一种强噪声背景下的语音端点检测算法。实验结果表明,相对于传统检测算法,WA-PSO-BP 提高了语音端点检测准确率,抗噪性、鲁棒性更好,具有更好的实际应用性。

参考文献:

- [1] Wu B F Y, Wang K C. Roust endpoint detection algorithm based on the adaptive band-partitioning spectral entropy in adverse environments[J]. IEEE Transactions on Speech and Audio Processing, 2005, 13(5): 762-775.
- [2] Cyril Plapous, Claude Marro, Pascal Scalart. Reliable a posteriori signal to- noise ratio features selection[C]// 2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics October, New Paltz, NY: IEEE, 2005: 16-19.
- [3] 乔峰,张雪英. 一种基于信息熵和神经网络的语音端点检测方法[J]. 太原理工大学学报, 2007, 12(38): 145-147.
- [4] BingFei Wu, KunChing Wang. Robust endpoint detection algorithm based on the adaptive band- partitioning spectral entropy in adverse environments [J]. IEEE Transactions on Speech and Audio Processing, 2005, 13(5): 762- 74.
- [5] 朴春俊, 马静霞, 徐鹏. 带噪语音端点检测方法研究[J]. 计算机应用, 2006, 26(1): 2685-2686.
- [6] 张铁威. 基于小波分析与神经网络的语音端点检测研究[D]. 大连:大连理工大学, 2008.

作者简介:

陈雪芳 女, (1978—), 硕士, 讲师. 研究方向为语音识别、数字图像处理等。