

小波分析和支持向量机相融合的语音端点检测算法

朱恒军 于泓博 王发智

(齐齐哈尔大学通信与电子工程学院 齐齐哈尔 161006)

摘 要 为了提高语音端点检测的适应性和鲁棒性,提出一种基于小波分析和支持向量机的语音端点检测算法。首先利用小波变换提取语音信号的特征量,然后将这些特征量作为支持向量机的输入进行训练和建模,最后判断出该信号的类别。仿真实验表明,相对于传统的语音端点检测算法,小波分析和支持向量机的检测算法提高了语音端点检测的正确率,有效降低了虚检率和漏检率,具有更好的适应性和鲁棒性,对不同信噪比的信号都有较好的检测能力。

关键词 小波分析,支持向量机,语音端点,特征提取

中图分类号 TP183 文献标识码 A

Speech Endpoints Detection Algorithm Based on Support Vector Machine and Wavelet Analysis

ZHU Heng-jun YU Hong-bo WANG Fa-zhi

(Communication and Electronic Engineering Institute, Qiqihar University, Qiqihar 161006, China)

Abstract In order to improve the adaptability and robustness of speech endpoint detection, this paper presented a algorithm for speech endpoint detection based on wavelet analysis and support vector machine. Firstly, the characteristic quantities of speech signals are obtained by the wavelet transformation. Then the input to support vector machine can be computed based on these characteristic quantities. Finally the signal's type can be determined. The simulation experiments results show that the proposed algorithm improves the detection rate, has better adaptability and robustness, and can detect signals with different SNR, compared with the traditional detection algorithms.

Keywords Wavelet analysis, Support vector machine, Speech Endpoints, Feature extraction

1 引言

语音信号是一种典型的非线性、非平稳随机信号,语音端点检测是语音识别系统至关重要的一步,可准确地从噪声中提取出语音信息,减少语音信号处理的运算量,从而大幅提高系统识别的准确率^[1]。

在语音信号处理系统过程,端点检测包括语音特征提取和检测器设计两个关键过程。语音特征提取方法主要有两类,一类是基于语音信号的时域处理,如短时相关分析、短时平均过零率和短时能量等检测算法^[2-4],该类算法在高信噪比下可以得到较准确的检测结果,但在低信噪比下,其准确率很低。在实际语音检测的应用中,往往很难保证有足够高的信噪比,容易将噪声当成语音。另一类是基于语音信号的频域处理,如谱熵法、频带方差和倒谱特征法等^[5],在好的条件下,该类检测方法可以取得好的检测结果,但在恶劣环境下,该类端点检测准确性仍然很差。各种检测方法均有各自的优点,可以进行互补,近些年多特征融合的语音端点检测算法不断出现,显著提高了低信噪比、噪声动态化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都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^[7]。近些年来,随着非线性科学的飞速发展,出现了基于神经网络和支持向量机(Support Vector Machines, SVM)的非线性语音

端点检测算法,相对于传统线性检测器,其提高了检测正确率,具有良好的区分性能^[8,9]。

为提高语音端点检测的正确率,提出一种小波分析和SVM相结合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,然后将这些特征量作为SVM的输入进行训练,SVM的参数采用粒子群算法进行优化,最后判断出该信号是语音信号或噪声。仿真对比结果表明,相对于传统检测算法,小波分析和SVM相结合的检测算法具有更好的语音检测效果,其适应性更广。

2 语音端点检测原理

语音端点检测又称语音活动检测(voice activity detection, VAD),通常是在复杂的噪声背景环境下用数字处理技术找出语音信号中的始点和终点位置,在信号流中分辨出语音信号和噪音,为后续信号处理提供必要的支持,其基本原理是通过语音和噪声对于相同参数所表现出的不同特征来区分两者的。小波分析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,因此采用小波分析来提取含噪语音信号的特征量。SVM是一种基于结构风险最小化原则的机器学习算法,较好地解决了小样本、非线性、维数灾等难题,不易产生类似神经网络的局部最优及过拟

到稿日期:2011-08-15 返修日期:2011-10-25 本文受齐齐哈尔市科技局科技攻关基金项目(GYGG-09007)资助。

朱恒军(1969—),男,硕士,副教授,主要研究方向为信号采集与信息处理、通信与信息系统, E-mail: hengjun_zhu@163.com。

合现象,泛化能力优异^[10]。利用 SVM 对经小波变换提取到的语音特征量进行训练和检测,以实现语音端点的检测。小波分析和 SVM 的语音端点检测原理框架如图 1 所示。

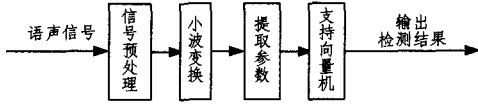


图1 语音端点检测原理框图

3 小波分析和支持向量机语音端点检测模型

3.1 语音信号的预处理

首先对语音数据进行预处理,包括采样、去噪、预加重、分帧、加窗、短时能量、平均过零率等工作。语音信号“四”在实验室环境下,在 Matlab 仿真平台上得到的波形、平均幅值和过零率如图 2 所示。

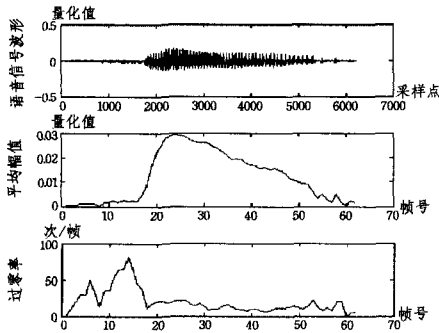


图2 语音“四”的波形、平均幅值和过零率

3.2 基于小波分析的语音特征量提取

3.2.1 信号的小波分解

小波分析的语音特征量提取是通过伸缩和平移一组小波基 $\varphi(t)$,使信号 $f(t)$ 的时频窗口随频率的变化而变化^[11]。小波变换的定义为:

$$WT_f(a, \tau) = \frac{1}{\sqrt{a}} \int f(t) \varphi\left(\frac{t-\tau}{a}\right) dt \quad (1)$$

式中, τ 表示时间移动, a 表示伸缩尺度。

首先将采集到的语音数据进行分帧处理,以每 16ms 为一帧,得到一组帧数据,采用低频时将窗函数拉伸、高频时将窗函数压缩的方法进行小波变换,然后对每帧信号 $f(n)$ 进行多分辨率分析(Mallat 快速算法),最后进行 5 层小波分解,分解公式如下所示:

$$\begin{aligned} f(n) &\rightarrow a_1(n) \rightarrow a_2(n) \rightarrow \dots \rightarrow a_5(n) \\ &\downarrow \quad \quad \quad \downarrow \quad \quad \quad \downarrow \quad \quad \quad \downarrow \\ d_1(n) & \quad d_2(n) \quad d_3(n) \quad \dots \quad d_5(n) \end{aligned} \quad (2)$$

首先将信号分解到高频空间 $d_1(n)$ 和低频空间 $a_1(n)$,再将 $a_1(n)$ 分解到高频空间 $d_2(n)$ 和低频空间 $a_2(n)$,然后继续分解 $a_2(n)$,一直分解到 5 层后得到 $d_1(n) \sim d_5(n)$ 、 $a_1(n) \sim a_5(n)$ 共 10 个不同的频率信号,根据小波多分辨率分析原理,其中 $d_1(n) \sim d_5(n)$ 、 $a_5(n)$ 可表征原信号所有的频率信号。

3.2.2 特征量提取

对语音信号频域的分析可知,语音信号在各个小波子带内的平均能量分布是不均匀的,信号的能量主要集中在低频子带内,即在小波子带 $a_5(n)$ 、 $d_5(n)$ 、 $d_4(n)$ 等内的平均能量要比 $d_6(n)$ 内的平均能量高。而噪声在各个子带内的平均能量分布比较均匀^[12]。计算各层小波子带系数的平均能量:

$$E_i = \frac{1}{N} \sum_{n=1}^N |s_i(n)| \quad (3)$$

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 为该小波子带每帧含有的小波系数的个数。在分别求出这 6 个小波子带的平均能量后,求其均值。

$$E_m = \frac{1}{6} \sum_{i=1}^6 E_i \quad (4)$$

然后计算各个小波子带平均能量的方差 σ^2 :

$$\sigma^2 = \frac{1}{6} \sum_{i=1}^6 (E_i - E_m)^2 \quad (5)$$

这样,就完成了对语音特征量的提取过程,每帧信号获得了 6 个小波子带 $d_1(n) \sim d_5(n)$ 、 $a_5(n)$ 的平均能量 $E_1 \sim E_6$,以及一个 E_m 和 σ^2 。这样 8 个特征量构成了该帧语音信号的特征矢量:

$$X(n) = [E_1, E_2, E_3, E_4, E_5, E_6, E_m, \sigma^2]^T \quad (6)$$

选择这 8 个量为特征量,是因为它们既可表现信号在不同时段能量,又可表现信号在各个频段的能量分布情况,还可以体现信号同一时刻不同频段的能量差异,完全表征了该帧信号的各种特点,通过它们可以完成对信号的语音端点检测。

3.3 SVM 的语音端点检测算法

3.3.1 SVM 分类算法

20 世纪 90 年代中期, Vapnik 等人提出支持向量机 (SVM), SVM 分类目标是寻找一个最优超平面,可以把需要分类的数据集合中的所有点分开,且使点尽可能地与该超平面距离最远,其分类示意图如图 3 所示^[13]。

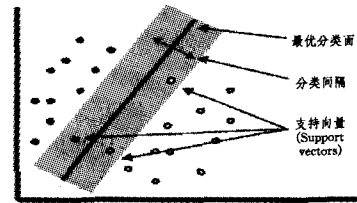


图3 SVM 的最优分类平面图

对两分类线性分类问题,给定数据集 (x_i, y_i) , $x_i \in R^n$, $y_i \in \{1, -1\}$, $i=1, 2, \dots, n$, x_i 表示分类的属性,即信号特征,其属于语音信号时,输出 $y_i=1$,否则 $y_i=-1$ 。SVM 的目标是根据风险最小化的原理,最优该超平面表示为:

$$y = w^T \varphi(x) + b \quad (7)$$

式中, w 为超平面法向量, b 为超平面偏移向量。

对非线性分类问题,需要将非线性分类问题转变成二次优化问题,即:

$$\min J(w, \xi) = \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n \xi_i \quad (8)$$

式中, c 为惩罚参数,对错分样本的惩罚程度进行控制。

式(8)的约束条件为:

$$\begin{aligned} y_i (w \cdot \Phi(x_i) + b) &\geq 1 - \xi_i \\ \xi_i &\geq 0, i=1, 2, \dots, n \end{aligned} \quad (9)$$

式中, $\xi = (\xi_1, \dots, \xi_n)^T$ 。

对于大样本的分类问题, SVM 学习速度慢,通过引入 Lagrange 乘子转化为对偶问题,然后对对偶问题进行求解,加快分类速度,得到 SVM 决策函数:

$$f(x) = \text{sign}(\sum_{i=1}^n a_i y_i (\varphi(x) \cdot \varphi(x_i)) + b) \quad (10)$$

式中, $sign$ 为符号函数, α_i 为 Lagrange 乘子。
通过核函数 $k(x_i, x)$ 来代替点积 $(\varphi(x) \cdot \varphi(x_i))$, 最后, SVM 的分类决策函数变为:

$$f(x) = sign(\sum_{i=1}^n \alpha_i y_i k(x_i, x) + b) \tag{11}$$

3.3.2 SVM 分类器的参数优化

SVM 对语音端点检测结果的优劣与其参数直接相关, 为此采用粒子群优化算法对支持 SVM 参数进行选择, 具体流程如图 4 所示。

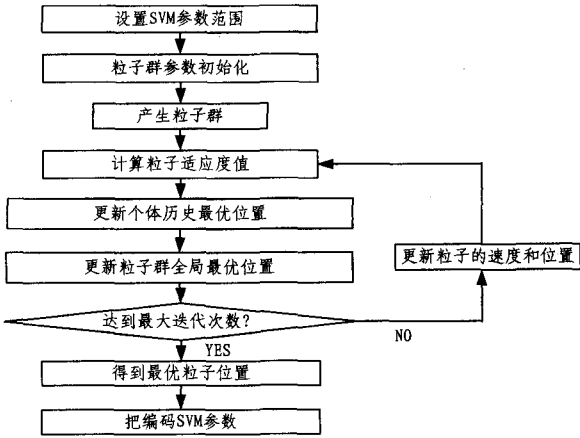


图 4 SVM 分类器参数的优化流程

3.3.3 SVM 的语音检测步骤

(1) 选择 SVM 模型。核函数选择径向基核函数 (Radial Basis Function, RBF), 设置参数 C 和径向基核函数宽度 σ 的范围。
(2) 任意选择一段信号作为 SVM 的训练样本, 该段信号中包含语音段和噪声, 首先人工标示出每帧信号类别, 若是语音帧则将其标示为 1, 非语音帧则标示为 -1, 然后对信号进行预处理, 并采用小波分析将信号分帧并提取出每帧信号的 8 个特征量, 每组特征量构成 SVM 的一组输入向量, 作为 SVM 的输入数据。对于每组输入数据, SVM 会产生 1 个输出结果, 其输出结果值为 -1 或 1。
(3) SVM 训练。将样本信号所提取的特征量作为 SVM 输入, 人工标示数据作为 SVM 的期望输出。SVM 训练要告知 SVM 当输入的信号特征量为某种形态时, 该段信号是语

音信号段, 因此该过程十分关键; 否则, 该段信号是噪声。由于 SVM 初始化参数随机设定, 最初的输出结果与期望结果间差异较大, 通过粒子群算法不断优化参数, 使 SVM 输出与理想输出值逐渐接近, 达到最大迭代次数后, 完成 SVM 训练。

(4) SVM 检验。在完成了 SVM 训练工作后, 仍以训练样本作为输入, 用训练好的 SVM 计算输出结果。将 SVM 输出结果与事先已标示好的信号语音帧情况进行对比, 检验 SVM 的训练效果, 结果一致, 则表示建立的 SVM 模型性能好, 否则认为 SVM 训练失败, 需要重新对 SVM 进行训练。

(5) 进行语音端点检测。取一段语音信号, 采用小波分析方法提取每帧信号的特征量, 将这些特征量作为 SVM 输入, SVM 输出结果作为对该帧信号的检测结果。

4 语音端点的仿真实验

4.1 数据来源

仿真实验数据来自互联网上下载得到的纯净语音数据文件, 以 .wav 的格式保存在计算机中, 具体如图 5 所示。采样率为 44100Hz、16 位量化、帧长 20ms, 结合 Cool Edit Pro 软件作为辅助的语音信号的分析工具, 对纯净语音样本的起点和终点进行手工标识, 并将其作为端点位置检查是否准确的标准。在 Matlab 平台上, 预处理后的语音信号采用人工加入的高斯白噪声, 信噪比 (SNR) 分别为 -5dB, 0dB, 10dB, 15dB。

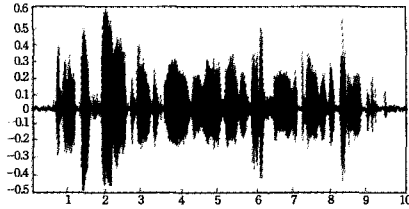


图 5 纯净语音

4.2 结果与分析

为了更好地对小波分析和 SVM 相结合的语音端点检测算法性能进行分析, 采用小波分析 + BP 神经网络、线性模型和短时能量法进行对比实验。采用正确率、虚检率和漏检率作为算法性能的评价标准。3 种算法的检测结果如表 1 所列。

表 1 几种算法对语音端点检测的识别率 (%)

SNR (db)	支持向量机			BP 神经网络			线性模型			短时能量法		
	正确率	虚检率	漏检率	正确率	虚检率	漏检率	正确率	虚检率	漏检率	正确率	虚检率	漏检率
-5	98.05	1.3	0.65	97.47	1.49	0.86	90.65	1.31	8.04	96.24	1.66	2.10
0	97.19	1.67	1.14	96.33	1.81	1.06	92.89	2.00	5.11	95.40	2.05	2.55
10	98.01	1.37	0.62	97.35	1.55	0.65	92.30	1.61	6.08	96.66	1.82	1.52
15	98.14	1.122	0.738	97.21	1.28	1.05	95.64	1.14	3.22	96.20	1.58	2.22
纯净音	99.75	0.25	0	99.27	0.29	0.51	97.80	0.27	1.93	98.02	0.73	1.25

从表 1 的结果可知, 在信噪比高的环境下, 所有的端点检测算法都具有良好的检测性能, 但随着噪声的加大, 对比语音端点检测算法, 性能下降更加明显, 有的已经不能达到实际运用的要求, 如线性检测模型。BP 神经网络的检验正确率要高于线性模型和短时能量法, 并且基于小波分析和 SVM 的检测算法好于基于 BP 神经网络算法。对比结果表明, 采用小波变换可以很好地提取信号特征量, 而采用非线性逼近能力强的 SVN 来检测语音端点, 可表现出更强的抗噪性, 且检测正确率有所提高, 漏检率降低。采用该算法来检测语音端点

可以获得比较满意的结果。
结束语 在噪声干扰不大的情况下, 传统语音端点检测算法可以正常工作, 但是当信噪比较小时, 其语音识别准确率急剧下降, 有时甚至无法正常工作, 把 SVM 和小波分析相结合, 并应用于语音端点检测中, 在不同的噪声环境下, 对算法性能进行仿真测试, 结果表明, 相对于传统检测算法, 小波分析和 SVM 相结合的算法准确率更高, 抗噪性更好, 有效提高了语音端点检测系统的稳定性, 更加适用于语音端点检测的
(下转第 265 页)

各类别邻域之间的距离来判定其所属类别。在 Urban、Washington 和 Indian Pine 数据集上的分类识别实验,说明了本文的方法能够较为有效地解决高光谱遥感影像分类的问题。在随机选取 4,6,8 个训练样本的情况下,本文方法的总体分类精度能够比其他算法提高 3.7%~8.5%。

但是本文的算法需要计算新样本与每个邻域的欧氏距离,计算复杂度偏大。而且当已知样本数量很少时,这种欧氏距离度量标准也可能会失效。如何解决这些问题,需要进一步研究。

参 考 文 献

[1] 梁亮,杨敏华,李英芳. 基于 ICA 与 SVM 算法的高光谱遥感影像分类[J]. 光谱学与光谱分析,2010,30(10):2724-2728

[2] 梁亮,刘志霄,杨敏华,等. 基于可见/近红外反射光谱的稻米品种与真伪鉴别[J]. 红外与毫米波学报,2009,28(5):353-356

[3] 王志辉,丁丽霞. 基于叶片高光谱特性分析的树种识别[J]. 光谱学与光谱分析,2010,30(7):1825-1829

[4] 余旭初,杨国鹏,冯伍法,等. 基于简约集支持向量机的高光谱影像分类[J]. 计算机科学,2010,37(11):268-270

[5] Turk M, Pentland A. Eigenfaces for recognition[J]. Journal of Cognitive Neuroscience, 1991, 3(1):71-86

[6] Belhumeur P, Hespanha J, Kriegman D. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1997, 20(7):711-720

[7] Ma Li, Crawford M M, Tian Jin-wen. Local manifold learning-based k-nearest-neighbor for hyperspectral image classification [J]. IEEE Trans. on Geoscience and Remote Sensing, 2010, 48(11):4099-4109

[8] Ma Li, Crawford M M, Tian Jin-wen. Generalized supervised lo-

cal tangent space alignment for hyperspectral image classification[J]. Electronics Letters, 2010, 46(7):497-498

[9] Roweis S, Saul L. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290(5500):2323-2326

[10] Tenenbaum J, Silva V D, Langford J C. A global geometric framework for nonlinear dimensionality reduction[J]. Science, 2000, 290(5500):2319-2323

[11] Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. Neural Computation, 2003, 15(6):1373-1396

[12] He Xiao-fei, Cai Deng, Yan Shui-cheng, et al. Neighborhood preserving embedding[C]// Proceedings of the 10th IEEE Int'l Conf. Computer Vision (ICCV'05), 2005. Beijing, 2005; 1208-1213

[13] He Xiao-fei, Yan Shui-cheng, Hu Yu-xiao, et al. Face recognition using laplacianfaces[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2005, 27(3):328-340

[14] Yan Shui-cheng, Xu Dong, Zhang Ben-yu, et al. Graph Embedding and Extensions: A General Framework for Dimensionality Reduction[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2007, 29(1):40-51

[15] 沈照庆,王建宾,陶建斌. 基于光谱相关性的 SVM 高光谱遥感影像分类[J]. 地理空间信息, 2009, 7(4):88-91

[16] 谭琨,杜培军. 基于径向基函数神经网络的高光谱遥感图像分类[J]. 光谱学与光谱分析, 2008, 28(9):2009-2013

[17] Fu Yun, Yan Shui-cheng, Huang T S. Classification and Feature Extraction by Simplexization[J]. IEEE Trans. on Information Forensics and Security, 2008, 3(1):91-100

[18] Fu Yun, Yan Shui-cheng, Huang T S. Discriminant Simplex Analysis[C]// Proceedings of the IEEE Conf. ICASSP, 2008. Las Vegas, NV, 2008:3333-3336

(上接第 246 页)
实际应用问题。

参 考 文 献

[1] 张雄伟,陈亮,杨吉斌. 现代语音处理技术及应用[M]. 北京:国防工业出版社,2003

[2] 刘华平,李听,徐柏龄,等. 语音信号端点检测方法综述及展望[J]. 计算机应用研究,2008,25(8):2278-2283

[3] 严剑峰,付宇卓. 一种新的基于信息熵的带噪语音端点检测方法[J]. 计算机仿真,2005,22(11):117-119

[4] 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(11):2685-2690

[5] Andreeut M, Este R A, Kauffman S A. Competitive optimization of compressed sensing[J]. Journal of Physics A: Mathematical and Theoretical, 2007(40):299-305

[6] Zhao H, Zhao X, Zhao K, et al. Voice activity detection based on distance entropy in noisy environment[C]// 5th International Joint Conference on INC, IMS and IDC. Seoul, Korea; IEEE Computer Society, 2009:1364-1367

[7] 乔峰. 基于信息熵和神经网络的语音端点检测算法研究[D]. 太

原:太原理工大学,2007

[8] 董恩清,赵鹤鸣,周亚同,等. 支持向量机在语音激活检测中的应用研究[J]. 通信学报,2003,24(3):70-77

[9] Giacobello D, Christensen M G, Murthi M N, et al. Retrieving Spame Patterns Using a Compressed Sensing Framework: Applications to Speech Coding Based on Spame Linear Prediction [J]. Signal Processing Letters, IEEE, 2010, 17(1):103-106

[10] 向昌盛,周子英,张林峰. 相空间重构和支持向量机参数联同步优化研究[J]. 湖南科学大学学报, 2010, 25(4):81-85

[11] 陈宝远,梁伟明. 基于小波分析的语音端点检测算法研究与仿真[J]. 哈尔滨理工大学学报, 2009(1):51-59

[12] Krishnamoorthy P, Prasanna S R M. Application of combined temporal and spectral processing methods for speaker recognition under noisy, reverberant or multi-speaker environment [J]. Indian Academy of Sciences, 2009, 10(34):729-754

[13] 隋海峰,曲武,钱文彬,等. 基于混合 SVM 方法的蛋白质二级结构预测算法[J]. 计算机科学, 2011, 38(10):169-174

[14] 郑继明,邢峰,吴渝,等. 基于小波变换和支持向量机的音频分类[J]. 重庆邮电大学学报:自然科学版, 2008, 20(2):212-216