

一种基于支持向量机的含噪语音的清/浊/静音分类的新方法

齐峰岩, 鲍长春

(北京工业大学电子信息与控制工程学院, 北京 100022)

摘要: 本文将支持向量机 (SVM) 方法应用于语音信号的清/浊/静音检测中, 提出并验证了一种在各种信噪比等级下将语音信号有效地分为清音、浊音和静音三类信号的新型分类算法. 首先, 在高信噪比情况下, 本文采用了 G. 729B VAD 中的四个差分参数作为 SVM 分类器的输入特征参数, 进行了静音分类的对比实验, 得到了优于 G. 729B VAD 和 BP 神经网络传统算法的实验结果, 说明引入这种机器学习方法做语音分类是可行的, 并分析讨论了在核函数不同的情况下支持向量机在实验中所表现出的性能. 其次, 又讨论了在低信噪比条件下, 如何通过含噪语音建立统计模型, 提取对噪音免疫的统计特征参数, 并给出了一种对时变背景噪声自适应的估计方法. 最后, 通过在不同噪音环境下的对比实验结果, 验证了本文所提出的算法在中低信噪比情况下的分类性能要优于其他传统算法.

关键词: 支持向量机; 统计学习; 统计信号处理; 模式识别; 语音编码

中图分类号: TN 912.3 **文献标识码:** A **文章编号:** 0372-2112 (2006) 04-0605-07

A Method for Voiced/Unvoiced/Silence Classification of Speech with Noise Using SVM

Q I Feng-yan, BAO Chang-chun

(School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100022 China)

Abstract A new method to voiced/unvoiced/silence of speech classification using Support Vector Machine (SVM) is proposed. This classifier can effectively classify speech frames into voiced frame, unvoiced frame and silence frame under various levels of signal noise ratio. Firstly, in high SNR, the VU/S classification is done by using the four difference characteristic parameters used in G. 729B VAD as SVM's input features. The comparison of experiment results shows that the proposed method outperforms other traditional methods (G. 729B VAD and BP network), which shows the SVM's classification method is available. And the performance of SVM for different kernel functions in the experiment was analyzed and discussed as well. Secondly, the paper also discusses the extraction of the statistical features which is immune to the background noise and the adaptive estimation method for the time-varying background noise in low SNR, which are analyzed by applying a statistical model. Lastly, the comparison experiment results in various noise environments under varying levels of SNR are given. According to the simulation results, the proposed method shows significantly better classification performances than the other traditional methods in middle and low SNR cases.

Key words support vector machine; statistic learning; statistical signal processing; pattern recognition; speech coding

1 引言

在语音信号处理中, 将语音信号可靠的分为清音、浊音和静音 (V/U/S) 是一件具有重要意义的事情, 例如在语

音编码^[1]中可以根据这三类信号的不同特点, 采用不同的处理方案以提高编码算法的效率和质量. 但是由于这三类信号的许多特征参数的取值动态范围在各类之间通常都是交叠的 (例如: 能量参数和过零率参数), 不可能通过提

取某一特征参数将其线性分开,而参数在受到噪音侵蚀的情况下就更为复杂了.在80年代以前一些经典的传统方法是通过提取某些特征参数,然后进行线性的处理和预定阈值来进行判断,阈值一般是靠人工经验来确定,方法简单,容易实现,但无法保证可靠精确的判断结果.这其中具有代表性的有1976年Atal和Rabiner提出的一种基于多个特征参数的V/U/S模式分类方法^[2],这种方法不依赖于基音检测,而是提取了过零率、能量、归一化自相关系数、第一个线性预测系数和归一化线性预测误差能量五个特征参数,它所使用的分类技术是一个贝叶斯决策过程.在这以后又产生了一些改进的和新型的判别方法如文献[3,4].80年代后随着人工神经网络的发展,许多学者将它引入到这一领域中,文献[5,6]分别介绍了一些应用不同特征参数和神经网络结构的判别方法.但传统的人工神经网络(如BP网络)方法在网络训练和网络设计上存在着训练速度慢、易陷入局部极小点和网络学习的推广性能差等缺陷,并且这种经验非线性方法在网络结构的选择和权重的初值设定仍需依靠使用者的工程经验,缺乏统一的数学理论基础.

90年代中期,Vapnik和他的AT&T Bell实验室小组在统计学习理论上发展出了一种新的模式识别方法——支持向量机^[7].在解决有限样本、非线性及高维模式识别问题中表现出许多特有的优势,避免了局部极小值和过拟合问题,提高了训练和测试的效率,在许多领域都得到了广泛的应用,成为机器学习领域研究的一个热点.

经典的支持向量机算法只能够进行二类分类,本文将两个二类分类支持向量机进行组合来实现语音信号的清音、浊音和静音分类,文献^[2-6]中提出的许多V/U/S判别方法都是对语音信号提取某些特征参数,然后根据判别函数将语音信号直接判为V/U/S三类信号中的一类,但由于许多特征参数在这三类信号中的取值范围都是交叠在一起的,同时分辨极为复杂.在本文中将其分为两步,每一步根据所分类别的特点选取不同的特征参数,使得判别更为准确,也易于实现,并且如果在第一步中判为静音则不进行第二步判决,这有利于最后的判别平滑算法,从而提高了整体判决的准确度和效率.

首先,在高信噪比情况下,本文采用ITU-T G.729B VAD中的四个差分参数作为SVM分类器的输入特征参数,进行了静音分类的对比实验,得到了优于G.729B VAD和BP神经网络传统算法的实验结果,并与文献[6]中提出的二级神经网络作了对比,说明此分类方法在语音分类中是有效的,并具有一定的性能优势.

但在实际应用中要求分类算法在低信噪比的情况下具有鲁棒性.在对含噪语音进行分类的研究中,从不同的分析角度出发,已提出了许多应用于实际的分类算法.例如,国际语音编码标准中的G.729B VAD和AMR VAD算法对背景噪音都具有一定的鲁棒性,但这些传统的分类算

法通常都是基于启发式的设计,这使得对分类参数很难进行优化.近些年来,一些文献提出了应用统计模型的方法进行VAD检测,通过提取对噪声免疫的判决参数来提升静音检测性能,其中,在文献[10]中给出了一个基于统计模型的似然比决策规则和一个自适应背景噪声的噪声参数估计器,文献[11]进一步完善了文献[10]的方法,并提出了一种基于一阶马尔可夫模型的VAD平滑算法.文献[12]中采用的是在去相关域中假设语音和噪声分别符合拉普拉斯和高斯分布的统计模型.文献[13]提出了一种利用统计高阶累积量作为判决参数的算法,以上算法各有特点,但它们所得出的实验结果均优于传统参数的分类方法,这说明在噪声环境下,许多传统的分类参数都已失效,因此需要借助统计模型的方法来提取对噪声具有鲁棒性的分类参数.在含噪语音的清/浊音分类研究中,文献[15]给出了一种基于模糊逻辑规则的对噪声具有鲁棒性的清/浊音分类器(FVD),文中的对比实验表明该算法在不同噪音环境下均优于美国联邦政府2.4kb/s语音编码标准方案中的清/浊判断算法.

本文对于低信噪比语音分类的情况,采用了统计参数(似然比和方差失真)作为SVM语音分类器的输入特征参数,并根据文献[10,11]相关内容提出了一种改良的自适应噪声参数估计方法和一个基于马尔可夫模型的清/浊/静音的分类平滑算法,最终的实验结果表明本文所提出的两步SVM清/浊/静音分类器算法在各种噪音环境下都保持了相对稳定的高准确率,而且在中低信噪比情况下,静音判决性能要远优于G.729B VAD,清/浊音判决性能则优于文献[15]中的FVD算法.

本文在第2节介绍SVM的基本原理,第3节介绍在安静环境下的分类算法描述,第4节介绍在低信噪比条件下的算法调整和分析,第5节讨论了含噪语音分类的实验数据和SVM的性能,最后是本文的结论.

2 支持向量机的基本原理

支持向量机方法是从线性可分情况下的最优超平面理论中提出来的,假定训练样本集是由两种类别数据组成:

$$(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l), \mathbf{x} \in \mathbf{R}^n, y \in \{+1, -1\} \quad (1)$$

这里的 y 为类别标号,如 $y_i = 1$ 表明 \mathbf{x}_i 为第一类数据,相反 $y_i = -1$ 则表明 \mathbf{x}_i 为第二类数据,对于 \mathbf{w} 和 b ,如存在一个超平面

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \quad (2)$$

可以将所有训练数据无错误地分开,即:

$$y_i = 1: \mathbf{w} \cdot \mathbf{x}_i + b \geq 1 \\ y_i = -1: \mathbf{w} \cdot \mathbf{x}_i + b \leq -1 \quad i = 1, \dots, l \quad (3)$$

并且离超平面最近的向量与超平面之间的距离是最大的,则称这个超平面是将这个训练集线性分开的最优超平面(最大间隔超平面).其中式(2)和(3)中的 $\mathbf{w} \in \mathbf{R}^n$, $b \in \mathbf{R}^1$

都进行了规范化,使得每类样本集中与分类超平面距离最近的数据点满足式(3)中的等式条件,这样分类间隔就等于 $2/\|w\|$,使其最大等价于使 $\frac{1}{2}\|w\|^2$ 最小,而使等号

$$f(x) = \text{sign}(w \cdot x + b) \quad i = 1, \dots, l \quad (4)$$

成立的那些样本称作支持向量.对式(3)可使用一种紧凑形式: 由统计学习理论可知,最优超平面能将两类无错误地分开是保证经验风险最小,而使分类距离最大实际上就是使推广性的界中的置信范围最小,从而使真实风险最小.这样得到的决策函数

$$f(x) = \text{sign}(w \cdot x + b) \quad (5)$$

其推广能力为最优.根据上面的讨论我们可以将最优超平面的求解问题归结为如下的二次规划问题:

$$\begin{aligned} \min \phi(w) &= \frac{1}{2} \|w\|^2 = \frac{1}{2} (w \cdot w) \\ \text{s.t. } & y_i (w \cdot x_i + b) \geq 1 \quad i = 1, \dots, l \end{aligned} \quad (6)$$

这个优化问题的解是由下面的拉格朗日函数的鞍点给出的^[7]:

$$L(w, b, \alpha) = \frac{1}{2} (w \cdot w) - \sum_{i=1}^l \alpha_i \{y_i [(w \cdot x_i) + b] - 1\} \quad (7)$$

其中, α_i 为拉格朗日乘子.把式(7)分别对 w 和 b 求偏导,并令它们等于 0 这样就可以将原问题转化为其较为简单的对偶问题:

$$\begin{aligned} \max Q(\alpha) &= \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \\ \text{s.t. } & \alpha_i \geq 0 \quad i = 1, \dots, l \\ & \sum_{i=1}^l y_i \alpha_i = 0 \end{aligned} \quad (8)$$

设 α_i^* 为解得的最优解,则

$$w^* = \sum_{i=1}^l \alpha_i^* y_i x_i \quad (9)$$

且根据 KKT(Karush-Kuhn-Tucker)条件可知:

$$\alpha_i \{y_i [(w \cdot x_i) + b] - 1\} = 0 \quad i = 1, \dots, l \quad (10)$$

因此,对大多数样本对应得 α_i 都为零, α_i 不为零则对应使式(4)等号成立的样本,即支持向量,它们只是全体样本中的一小部分.并根据此式取任意支持向量可解出 b^* ,由此可得到基于最优超平面的分类规则的指示函数:

$$f(x) = \text{sign} \left(\sum_{\text{支持向量}} y_i \alpha_i^* (x \cdot x_i) + b^* \right) \quad (11)$$

上面讨论的是线性可分的情况,然而实际的情况往往是不可分的,对这种情况需对原来最优化问题作些修正,在式(4)中引入非负变量 $\xi \geq 0$ 变成:

$$y_i (w \cdot x_i + b) \geq 1 - \xi \quad (12)$$

而式(6)就变为如下的二次规划问题:

$$\begin{aligned} \min \phi(w) &= \frac{1}{2} (w \cdot w) + C \left(\sum_{i=1}^l \xi_i \right) \\ \text{s.t. } & y_i (w \cdot x_i + b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \quad i = 1, \dots, l \end{aligned} \quad (13)$$

其中 C 为惩罚参数,它实际表征了对错分样本的惩罚程度, C 越大表示对错误分类的惩罚越大.求解这个二次优化问题的技术与在线性可分情况下几乎相同,只是式(8)中的条件变为:

$$0 \leq \alpha_i \leq C, \quad i = 1, \dots, l \quad (14)$$

在上面讨论的最终分类指示函数(11)和求解过程式(8)中,可以发现它们只涉及到待分样本与训练样本中的支持向量的内积运算 $(x_i \cdot x)$ 或训练样本之间的内积运算 $(x_i \cdot x_j)$,可见,要解决一个特征空间中的最优线性分类问题,只需知道这个空间中的内积运算即可.如果一个问题在其定义的空间中是线性不可分的,这时可以考虑通过某种事先选择的非线性映射,把问题映射到一个新的特征空间,这个空间一般比原空间维数要增加,但却可以用线性判别函数实现原空间中的非线性判别函数.考虑在 Hilbert 空间中内积的一个一般表达:

$$(z_i \cdot z) = K(x, x_i) \quad (15)$$

其中 z 是输入空间中的向量 x 在特征空间中的像.根据 Hilbert-Schmidt 理论, $K(x, x_i)$ 可以是满足 Mercer 条件的任意对称函数.如果用 $K(x, x_i)$ 代替式(8)和(11)中的点积,此时在输入空间中非线性的决策函数为:

$$f(x) = \text{sign} \left(\sum_{\text{支持向量}} y_i \alpha_i^* K(x_i \cdot x) + b^* \right) \quad (16)$$

这就相当于把原空间变换到了某一新的特征空间去构造最优分类超平面,其中 $K(x, x_i)$ 一般被称作核函数,算法的其他条件不变,这就是支持向量机(如图 1).

从最终的判别函数可看出,此算法的计算复杂度取决于支持向量的个数,而不是变换空间的维数.根据统计学习理论可知,支持向量机的推广性也是与变换空间的维数无关的,只要能够适当地选择一种内积函数,构造一个支持向量数相对较少的最优分类超平面,则就可以得到较好的推广性.这与传统方法有着截然相反的思路,传统的方法试图经过处理(特征选择和特征变换)将原输入空间降维,而支持向量机方法是将输入空间升维,以求在高维空间中将原问题变成线性可分或接近线性可分,这种方法之所以可行是因为升维后算法复杂度并不随维数的增加而增加,在高维空间中的推广能力也并不受维数的影响,并可以很好地避免“维数灾难”问题.

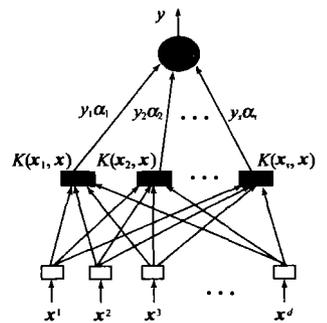


图 1 支持向量机示意图

3 在安静环境下的算法描述

3.1 算法的流程图

本算法采用了两个二类支持向量机进行组合来实现语音的清音、浊音和静音三类信号的分离,算法流程如图 2 首先输入语音经过一个截止频率为 140Hz 的二阶 IIR 高

通滤波器, 滤除不必要的低频干扰, 然后提取四个差分参数 (参数定义在下节中给出), 将参数输入到第一个已训练好的支持向量机分类器, 先对语音信号进行有声无声判决, 如果被判为静音, 就直接做最后的判决平滑, 如果判为有声再提取相应参数做清浊判决, 然后再做判决平滑。两步判决中的参数提取具有一定的承接性, 最后判决结果通过判决平滑去除判决中存在的“孤立点”, 判决平滑采用了基于一阶马尔可夫模型的平滑算法, 这种两步判决算法有利于提高算法效率和判决的准确度, 避免了统一判决的复杂性。

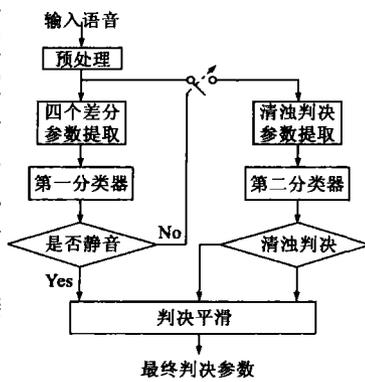


图2 算法流程图

然后再做判决平滑。两步判决中的参数提取具有一定的承接性, 最后判决结果通过判决平滑去除判决中存在的“孤立点”, 判决平滑采用了基于一阶马尔可夫模型的平滑算法, 这种两步判决算法有利于提高算法效率和判决的准确度, 避免了统一判决的复杂性。

3.2 特征参数及算法实现

提取参数的原则是特征参数要易于提取, 实现简单, 对不同模式的分类可靠有效, 参数的取值范围在待分类别中的交叠较少, 并要求特征参数要从不同角度来描述样本特性, 这样便于识别。下面给出了两个分类器所涉及到的八个特征参数。

3.2.1 全带能量 (E_f)

$$E_f = 10 \times \log_{10} \left[\frac{1}{N} R(0) \right] \quad (17)$$

其中, $R(0)$ 为信号的短时自相关函数。

3.2.2 低带能量 (E_l)

将语音信号通过一个 800kHz 的低通滤波器, 该滤波器采用的是一个 5 阶椭圆 IR 滤波器, 通带截止频率为 800Hz, 阻带起始频率为 1200Hz, 通带波动小于 0.25dB, 阻带波动小于 50dB。将滤波后的信号利用式 (17) 计算出低带能量^[1]。

3.2.3 过零率 (ZC)

$$ZC = \frac{1}{2M} \sum_{i=0}^{M-1} | \text{sgn}[x(i)] - \text{sgn}[x(i-1)] | \quad (18)$$

3.2.4 全带能量差 (ΔE_f)

$$\Delta E_f = \bar{E}_f - E_f \quad (19)$$

3.2.5 低带能量差 (ΔE_l)

$$\Delta E_l = \bar{E}_l - E_l \quad (20)$$

3.2.6 过零率差 (ΔZC)

$$\Delta ZC = \bar{ZC} - ZC \quad (21)$$

3.2.7 谱失真 (ΔS)

$$\Delta S = \sum_{i=1}^p (LSF_i - \overline{LSF_i})^2 \quad (22)$$

3.2.8 周期性水平 (Z_{period})

$$Z_{\text{period}} = \rho_{\text{max}} + \rho_{\text{avr}} \quad (23)$$

$$\text{其中, } \rho(t) = \frac{\sum_{n=0}^{N-1} s(n)s(n-t)}{\sqrt{\sum_{n=0}^{N-1} s^2(n) \sum_{n=0}^{N-1} s^2(n-t)}} \quad \rho_{\text{max}} \text{ 为在基音范}$$

围中 ρ 的最大峰值, ρ_{avr} 为 ρ 的前三个峰值的平均值。

第一分类器用到的是全带能量差 (ΔE_f)、低带能量差 (ΔE_l)、过零率差 (ΔZC)、谱失真 (ΔS) 四个差分参数^[8]。第二分类器用到的是全带能量 (E_f)、低带能量 (E_l)、过零率 (ZC)、周期性水平 (Z_{period})^[11] 四个特征参数。针对不同的分类器对每帧语音分别提取特征参数, 再将四维特征参数作为输入向量进行训练和测试, 支持向量机算法的实现是采用 Joachims 所提供的 SVM_{lib} 的最优化算法, 详细内容参阅文献^[9]。代码编写参照了 Stefan Ruppinger's smySVM 软件包所提供的代码。

3.3 实验结果与分析

算法的实验数据是采样率为 8KHz, 16bit 的 PCM 格式的数字语音信号。每帧帧长为 10ms 即 80 个样点。训练样本为不同年龄段的男女各四名的不同语句的发音 (取自北京工业大学语音与音频信号处理实验室语音数据库), 训练样本共 1780 帧, 约 18 秒, 其中语音占 42.68%, 由 34.27% 的浊音和 8.41% 的清音组成。训练样本的分类是人工通过其时域波形特性、频域频谱特性和它实际对应的音素的特性综合判定的。检测样本取自训练样本集外的不同语句的语音样本, 共 1236 帧, 约 12 秒, 其中语音占 37.59%, 由 30.43% 的浊音和 7.14% 的清音组成。本文算法在有声无声实验中与 G.729B VAD 结果作了性能对比, 在 G.729B VAD 中组合了 14 个线性分类面来做有声无声检测, 而 SVM 分类器是根据典型小样本训练集自动寻求一个非线性的最优分类面, 实验表明 SVM 方法的性能要优于 G.729B VAD; 本文算法在清浊静音分类实验中与 BP 网络以及文献^[6]中的二级神经网络作了性能对比。下表 1 给出了这些对比实验结果, 实验中的统计结果表明此算法要优于传统的分类算法, 可见应用支持向量机的方法不但是可行的, 并且在语音分类中具有一定的性能优势。

表 1 支持向量机与 BP 网络、G.729B VAD 性能对比

算法	有声无声判决	清音浊音判决
	准确度 /%	准确度 /%
G.729B VAD	97.81	无
BP 网络	90.16	96.34
二级神经网络 ^[6]	98.25	98.49
支持向量机	98.95	98.68

图 3 给出了一个直观的实验结果图, 此语音是从测试样本中任意选出的一句女声语句 (“塞翁失马”)。从图中可以看出本文算法不但能够将能量较低的清擦音 s 与静音和浊音分开, 而且也能够准确地将能量较高的清擦音 sh 从浊音中区分出来。

在实验中发现选取适当的核函数及其参数对提高

SVM 分类器性能是十分必要的。表 2 给出了在相同实验条件下选用不同的核函数及其参数进行清浊判决的实验结果。

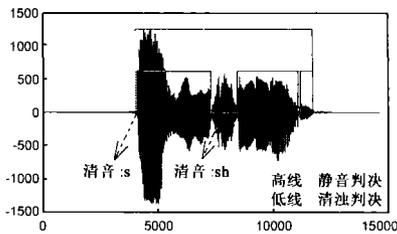


图 3 实验结果

从表中可以看出采用径向基核函数的实验结果最好, 这说明在本试验中并不是核函数越复杂, 实验效果就越好, 而是取决于适当地选择一种核函数使所构造的分类超平面在新的映射空间中能更准确的进行分类。但如何根据实际情况, 选择合适的核函数, 还是这一领域今后有待研究的问题。

4 在低信噪比条件下的改进算法

本文在讨论含噪语音时, 首先假设背景噪声为与语音信号不相关的加性噪声, 例如, 在语音通信中经常遇到的时变嘈杂人声 (Babble) 背景噪声 (取自 NOISEX_92 噪声库), 在这种噪声的低信噪比情况下如仍采用 G. 729B VAD 中的四个差分参数来做有声无声检测, 性能就会急剧下降, 因为部分参数受噪声影响已很难区分开有声无声了。在这种情况下需要找到一些对背景噪声具有一定免疫力的特征参数。因此, 本文借助了统计模型的方法在频域中

提取似然比参数作为 SVM 分类器的输入特征参数, 来提高分类算法对噪声的鲁棒性。为了适应时变噪声, 还必须设计一个能够准确跟踪时变的背景噪声

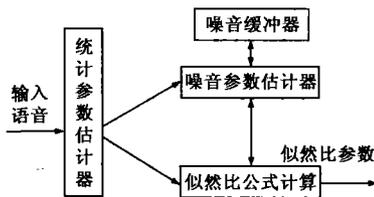


图 4 似然比参数产生框图

的噪声估计器, 图 4 给出了似然比参数的产生流程图。

表 2 支持向量机不同核函数性能对比

核函数	内积函数参数	清浊音判决准确度 /%
多项式内积	$d = 4$	97.4
径向基函数内积	$\sigma^2 = 0.3$	98.2
Sigmoid 内积	$a = 0.25, b = 1$	96.8

注: (1) 多项式内积函数为 $K(\mathbf{x}, \mathbf{x}_i) = [(\mathbf{x} \cdot \mathbf{x}_i) + 1]^d$

(2) 径向基内积函数为 $K(\mathbf{x}, \mathbf{x}_i) = \exp\left\{-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{\sigma^2}\right\}$

(3) Sigmoid 内积函数为 $K(\mathbf{x}, \mathbf{x}_i) = \tanh[a(\mathbf{x} \cdot \mathbf{x}_i) - b]$

(表中参数仅供参考, 结果未为做判决平滑处理)

4.1 似然比参数描述

首先, 假设语音与加性背景噪声是互不相关的, 含噪语音 $y(n)$ 的帧长为 T , 对其做离散傅立叶变换 (DFT), 用 Y, N 和 S 分别表示含噪语音、干净语音和背景噪声的 T 维 DFT 系数矢量。在这里应用高斯统计模型, 认为每一维的 DFT 系数是渐近独立的高斯随机变量^[14]。如针对有声无声

检测问题, 则有如下两种假设:

$$H_0: \text{无语音: } Y = N \quad (24)$$

$$H_1: \text{有语音: } Y = N + S \quad (25)$$

这样就可以得到在关于 H_0 和 H_1 的联合条件概率密度函数:

$$p(Y|H_0) = \prod_{k=0}^{T-1} \frac{1}{\sigma^2(k)} \exp\left\{-\frac{|Y_k|^2}{\sigma^2(k)}\right\} \quad (26)$$

$$p(Y|H_1) = \prod_{k=0}^{T-1} \frac{1}{\sigma^2(k) + \lambda^2(k)} \exp\left\{-\frac{|Y_k|^2}{\sigma^2(k) + \lambda^2(k)}\right\} \quad (27)$$

这里的 $\sigma^2(k)$ 和 $\lambda^2(k)$ 分别表示为噪声和语音 DFT 谱第 K 维成分的方差, 将此式代入似然比公式可得相应的似然比为:

$$\log \Lambda = \frac{1}{L} \log \frac{p(Y|H_1)}{p(Y|H_0)} = \frac{1}{L} \sum_{k=0}^{T-1} \left\{ \frac{\lambda_k^2 \xi_k}{1 + \lambda_k^2 \xi_k} - \log(1 + \xi_k) \right\} \quad (28)$$

其中, $\xi_k = \lambda^2(k) / \sigma^2(k)$ 和 $\lambda_k^2 = |Y_k|^2 / \lambda^2(k)$, 它们分别被称作先验信噪比和后验信噪比^[14]。在这里假设 $\lambda^2(k)$ 已经知道, 因为它可以通过后面的噪声统计估计器得到, 那么现在就必须估计另一个未知参数 ξ_k , 这里采用文献 [14] 中讨论的最大似然估计方法, 得到如下结论:

$$\xi_k(n) = \begin{cases} \bar{\lambda}_k(n) - 1, & \bar{\lambda}_k(n) - 1 \geq 0 \\ 0, & \text{其它} \end{cases} \quad (29)$$

其中, $\bar{\lambda}_k(n) = \alpha \bar{\lambda}_k(n-1) + (1-\alpha) \frac{\lambda_k^2(n)}{\beta}$, $0 \leq \alpha < 1, \beta \geq 1$

α, β 表示修正因子, 在试验中可以根据噪声的不同特点进行自适应调节, 将式 (29) 代入到式 (28), 就可以得到一个最终的似然比表达式:

$$\log \Lambda = \frac{1}{L} \sum_{k=0}^{T-1} \{ \bar{\lambda}_k - \log \bar{\lambda}_k - 1 \} \quad (30)$$

图 5 出示了此参数在车辆背景噪声下的分类效果, 这种噪声周期性较强, 能量主要集中在低频段, 噪声源来自 NOISEX-92 噪声库, 信号源来自一段典型汉语语音, 图 5 展示了似然比参数性能 (背景噪声为车辆噪声, SNR=10dB); (a) 似然比的对数参数 (b) 原始语音 (c) 加噪语音中数据已归一化。

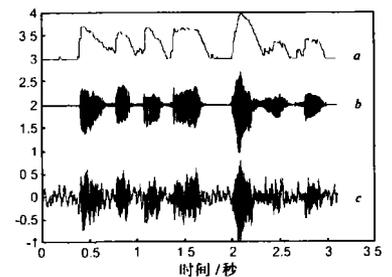


图 5 似然比参数性能 (背景噪声为车辆噪声, SNR=10dB); (a) 似然比的对数参数 (b) 原始语音 (c) 加噪语音

4.2 自适应噪声的估计

在上述似然比参数讨论中曾遗留下一个问题就是假设背景噪声方差 $\lambda^2(k)$ 为已知, 而此参数的估计是否准确对似然比参数的有效性至关重要, 必须能够自适应背景噪声的变化。在 G. 729B VAD^[8] 中也对背景噪声估计采取了一些自适应的启发式算法, 首先利用前 32 帧信号的信息升

级最初的噪声平均参数,然后仅当帧能量小于平均噪声能量与某一常数的和时,才应用一阶自回归系统进行更新.这种方法存在着一些固有缺陷,对时变的噪声源(例如Babble)分类效果并不理想.在文献[10]中提出了一种应用软判决信息对噪声谱进行自适应估计的方法,对时变噪声的估计十分有效,估计器是通过一阶无限冲激响应(IR)滤波来实现的,具体表达如式(31),具体推导详见文献[10]的3.1节.

$$\hat{\lambda}^{(m)}(k) = \frac{1}{1 + \varepsilon\Lambda^{(m)}} |X_k^{(m)}|^2 + \frac{\varepsilon\Lambda^{(m)}}{1 + \varepsilon\Lambda^{(m)}} \hat{\lambda}^{(m-1)}(k) \quad (31)$$

其中, $\varepsilon = \frac{p(H_{1,k})}{p(H_{0,k})}$, $\Lambda^{(m)}$ 为当前帧的似然比, $\hat{\lambda}^{(m-1)}(k)$ 是前一帧的估计.本文改进了此算法,使用一个噪声缓冲器来判定背景噪声的变化速度,自适应调节噪声谱估计器的更新参数,表达式如下:

$$\hat{\lambda}^{(m)}(k) = t^{(m)}(k) + (1 - t^{(m)}(k)) \hat{\lambda}^{(m-1)}(k) \quad (32)$$

在实验中发现对典型的时变噪声Babble(嘈杂人声)的估计十分有效,具有一定的自适应性和平滑性.从第5节的试验结果也可以看到这一点.

4.3 含噪语音的算法流程

首先在特征参数选用上除了上面提到的似然比参数、能量差参数、谱失真参数,本文还根据语音与噪声的方差统计特性的不同,提取了方差失真参数,用当前帧的方差与估计的背景噪声的方差相减,因为对于许多背景噪声符合高斯分布的情况,噪声方差较稳定而语音帧的方差却与其相差较大,所以在语音帧时此参数较小,而在语音帧则较大.现将以上提及的四个特征参数作

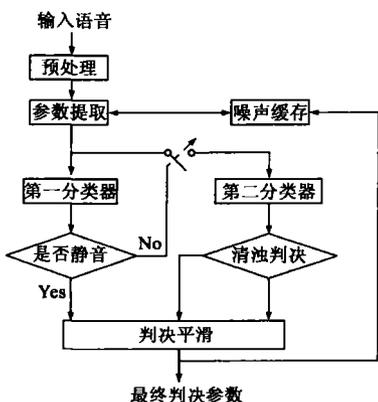


图6 含噪语音判别算法流程图

为一个四维输入矢量,经过两步SVM分类器判决,产生初始的输出结果,再经过一个基于马尔科夫统计模型的平滑算法得到最后的判决结果,并以此结果为依据来更新噪声缓冲器,最终算法流程如图6

5 实验结果

试验中采用的语音数据与第3节中的试验数据是相同的,背景噪声数据来自NOISEX-92噪声库,首先将噪声库数据调整到8k/s采样率,再

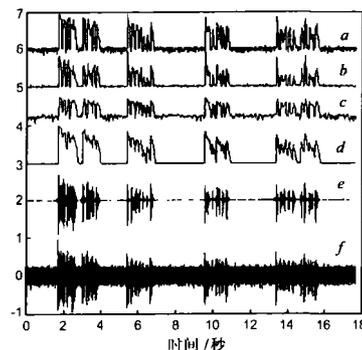


图7 四个特征参数变化曲线(背景噪声为SNR=0dB的白噪声)(a)能量差(b)谱失真(c)方差失真(d)似然比(e)原始语音(f)含噪语音

进行加性混合,得到在不同噪声环境下不同信噪比等级的含噪实验语音.首先,图7给出了四个参数在信噪比为零的白噪声条件下的性能.从图中可以看出这四个参数在白噪声条件下都具有较强的区分能力.从四个参数的对比可以看出似然比参数和谱失真参数要略优于另外两个参数,但在嘈杂人声(Babble)环境下,谱失真参数就失去了一定的有效性,而在车辆噪声环境下,能量差参数的性能就会变差,试验中发现虽然在不同的噪声环境下各个参数的分类能力互有优劣,但是SVM分类器的判别输出却不会因某一个特征参数性能变坏而使总的判决准确率急剧下降.本文在白噪声、车辆噪声和嘈杂人声三种典型噪声条件下对所提算法的性能做了实验验证.图8至图10出示了在这三种噪声的不同信噪比条件下的性能对比图,图中的四类曲线分别表示本文算法的静音判决准确度,清浊判决准确度以及G.729B VAD的判决准确度和文献[15]中给出的FVD清浊判决准确度结果.从实验数据曲线可以看出,本文所提出的算法在各种噪声环境下都保持了相对稳定的高准确率,在中低信噪比情况下静音判决性能要远远优于G.729B VAD,并且在白噪声和嘈杂人声情况下清浊判

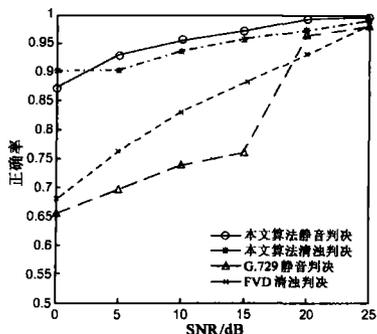


图8 背景噪声为白(White)噪声的分类性能对比图

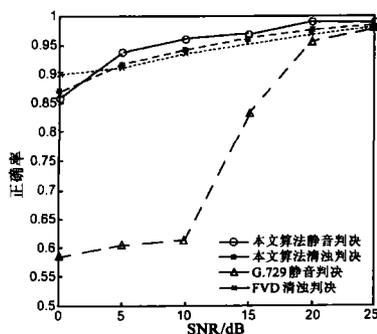


图9 背景噪声为车辆(Vehicle)噪声的分类性能对比图

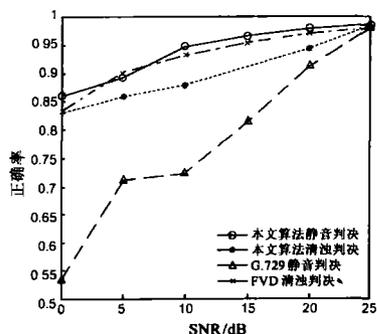


图10 背景噪声为嘈杂人声(Babble)的分类性能对比图

决的性能也明显优于 FVD 算法的性能。

6 结论

从整个实验结果可以看出,这种两步 SVM 清/浊/静音语音分类器在本实验中的性能要优于传统的模式分类算法,而且在噪音环境下,SVM 分类器也不会因为某一维特征参数变差导致最后输出性能急剧下降。这种机器学习的方法与传统方法相比,避免了人工预设经验阈值的局限性,并可以通过较小的训练样本集获得相当好的分类结果。支持向量机在做判决时它先通过核函数进行非线性变换将输入空间转到一个高维空间,然后在这个新空间中用凸二次规划理论求出最优线性分类超平面,避免了局部最小值的问题,SVM 构造的复杂程度取决于支持向量的数目,而不是特征空间的维数,很好地避免了“维数灾难”问题,因而在训练效率、测试效率、克服过拟合、算法参数调节和推广性等方面支持向量机都有着优异表现。但这种基于统计学习理论的学习算法还处在不断发展完善中,在快速求解算法、多类别分类、参数的自适应选择、先验知识的利用以及理论的进一步完善等方面还有待深入的研究和探讨。而在低信噪比噪声环境下,如何结合先验统计知识更好地自适应估计噪声以及寻求对噪声更有效的特征参数等方面也是今后提高 SVM 语音分类算法性能的研究重点。

总之,本文提出了一种新的基于 SVM 的清/浊/静音分类算法,在不同背景噪声环境下可有效地区分开这三类信号,为那些需要对语音信号进行分类处理的应用方案提供了一种可供参考的对噪声具有鲁棒性的有效算法。

参考文献:

- [1] 鲍长春.低比特率数字语音编码基础[M].北京:北京工业大学出版社,2001
- [2] B A ta L Rabiner A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition[J]. IEEE Transactions on Acoustics Speech and Signal Processing 1976 24(3): 201-212
- [3] L Rabiner M Sambur Application of an LPC distance measure to the voiced-unvoiced-silence detection problem[J]. IEEE Transactions on Acoustics Speech and Signal Processing 1977 25(4): 338-343
- [4] B Cox L M Timothy Nonparametric rank-order statistics applied to robust voiced-unvoiced-silence classification[J]. IEEE Transactions on Acoustics Speech and Signal Processing 1980 28(5): 550-561
- [5] Y Qi B R Hunt Voiced-unvoiced-silence classification of speech using hybrid features and a network classifier[J]. IEEE Transactions on Speech and Audio Processing ,

1993, 1(2): 250-255

- [6] R Ahn W H Holmes Voiced/unvoiced/silence classification of speech using 2-Stage neural networks with delayed decision input[A]. B Boashash et al Proc ISSPA 96[C]. Brisbane Australia Queensland University of Technology, 1996 389-390
- [7] Vladimir N Vapnik 统计学习理论的本质[M].张学工译.北京:清华大学出版社,2000
- [8] ITU-T Rec G. 729-1996 ANNEX B, A silence compression scheme for G. 729 optimized for terminals conforming to Recommendation V. 70[S].
- [9] B Schokopf C Burges A Smola Advances in Kernel Methods-Support Vector Learning[M]. USA: MIT Press, 1999 41-56
- [10] Jongseo Sohn Wonyong Sung A voice activity detector employing soft decision based noise spectrum adaptation[A]. Proc ICASSP'98[C]. Seattle Washington IEEE, 1998 365-368
- [11] Jongseo Sohn Nam Soo Kim, Wonyong Sung A statistical model-based voice activity detection[J]. IEEE Signal Processing Letters 1999, 6(1): 1-3
- [12] S Gazor Wei Zhang A soft voice activity detector based on a Laplacian-Gaussian model[J]. IEEE Transactions on Speech and Audio Processing 2003 11(5): 498-505
- [13] E Namer R Goubian SM Almoud Robust voice activity detection using higher-order statistics in the LPC residual domain[J]. IEEE Transactions on Speech and Audio Processing, 2001, 9(3): 217-231
- [14] Y Ephraim, D Mahk Speech enhancement using a minimum mean square error short-time spectral amplitude estimator[J]. IEEE Transactions on Acoustics Speech and Signal Processing 1984 32(6): 1109-1121.
- [15] F Beritelli S Casale Robust voiced/unvoiced speech classification using fuzzy rules[A]. 1997 IEEE Workshop on Speech Coding For Telecommunications Proceedings[C]. Pennsylvania USA: Voxware Inc, 1997 5-6

作者简介:



齐峰岩 男,1975年10月出生,吉林辽源,博士研究生,研究方向为机器学习在语音信号处理中的应用、语音编码等。

E-mail qifengyan@emails.bjtu.edu.cn