



基于支持向量机的非特定人孤立数字语音识别

陈少杰¹, 孙 敏², 柳映辉¹

(1. 天津科技大学计算机科学技术与信息工程学院, 天津 300222; 2. 中国农业银行菏泽市分行, 菏泽 274000)

摘 要: 为了识别一组非特定人、不连续的数字语音信号, 本文提出了一种基于支持向量机理论的语音信号识别算法. 具体过程主要包括训练过程和识别过程. 其中训练过程为: 先使用预先建立起来的语音库对选定的支持向量机进行训练, 得到一组与该语音信号相关的支持向量; 在识别过程中, 首先获取被测语音信号, 并根据 MFCC 理论提取特征向量, 然后使用训练后的支持向量机进行识别. 此外, 还提出使用短时区域能量谱的方法对语音信号进行端点检测. 结果表明, 与目前流行的隐马尔可夫算法比较, 本文算法具有识别速度快、准确率高等优点.

关键词: MFCC; 支持向量机; 语音识别; HMMD

中图分类号: TN912.3

文献标志码: A

文章编号: 1672-6510(2009)02-0059-04

Recognition of Discontinuous Numeric Speech by Nonspecific Person Using Support Vector Machines

CHEN Shao-jie¹, SUN Min², LIU Yin-hui¹

(1. College of Computer Science & Information Engineering, Tianjin University of Science & Technology, Tianjin 300222, China; 2. Heze Branch, Agriculture Bank of China, Heze 274000, China)

Abstract: To distinguish a group of nonspecial discontinuous numeric speech signals (nDNSS), a algorithm using Support Vector Machines (SVM) was proposed. The whole process included two main steps of training and recognition. SVM was selected in the training by using a preestablished nDNSS database. During recognition process, an eigenvector of a nDNSS was drawn out primarily using MFCC theory and then taken into the selected SVM to be recognized as input parameters. Furthermore, an algorithm for end-point detection of speech signal based on short-term-region-energy spectrum (STRES) was presented for the first time. Experiments show that the proposed algorithm can achieve better performance than the existing popular method (HMMD) on rapidity, accuracy and robustness.

Keywords: MFCC; support vector machines; voice recognition; HMMD

语音识别是语音处理技术中重要组成部分, 其最终目标就是让机器能够“听懂”人类语言. 自 20 世纪 50 年代世界上第一个语音识别系统——Audry^[1]诞生至今, 语音识别技术一直是研究热点之一. 理论上, 从倒谱 (Cepstrum) 分析到线性预测编码 (Linear Predict Coding, LPC), 从隐马尔可夫模型 (Hidden Markov Model, HMMD)^[2]到小波变换 (Wavelet Transform, WT)^[3], 语音识别理论得到长足发展; 应用上, 也从最初仅仅能识别有限个孤立语音信号的 Audry 系统, 发

展到当前能够识别所有基本语音且识别率在 90% 以上的 ViaVoice 系统^[4].

尽管建立在隐马尔可夫模型基础上的语音识别取得了一定成功, 但马尔可夫理论是在假设各个语音信号之间相互独立的基础上的^[5]. 对于连续的语音信号, 尤其是在语音信号间出现语音混叠情况下, 很难满足这一要求. 出现于 20 世纪 90 年代的支持向量机 (Support Vector Machine, SVM)^[6], 恰恰克服了这一不足. 该算法以核特征空间为核心, 巧妙地结合了

收稿日期: 2008-08-29; 修回日期: 2008-12-11

基金项目: 天津科技大学自然科学基金资助项目 (20060221)

作者简介: 陈少杰 (1972—), 男, 江苏徐州人, 讲师.

泛化理论和最优化理论,它使用 Karush-Kuhn-Tucker 和 Mercer 条件,使得求解最优化问题成为凸问题,从而成功解决了优化问题求解过程产生局部最小的不足,并成为有效的学习算法^[7]. 支持向量机的出现,极大地推动了语音识别的进一步发展,本文正是在此背景下进行的一次探究.

1 支持向量机原理

现有一个两类训练样本集, $D=\{(x_1, y_1), \dots, (x_l, y_l)\}$, $x \in \mathbf{R}^n, y \in \{-1, 1\}$, 为了把该样本集中的两类样本有效地区分开,只要找到一个超平面 $(w, x) + b = 0$, 使得距离超平面最近的两类点间隔最大. 此时的分类间隔为 $2\|w\|$, 为了寻求最大分类间隔,可以将该问题转化为如下二次规划问题:

$$\begin{cases} \min \frac{1}{2} \|w\|^2 \\ \text{s.t. } y_i [\langle w, x_i \rangle + b] \geq 1, \quad i=1, \dots, l \end{cases} \quad (1)$$

式(1)的最优解可以通过求解拉格朗日函数的鞍点得到.

$$\Phi(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i (y_i [\langle w, x_i \rangle + b] - 1) \quad (2)$$

式中: α 拉格朗日乘子. 依据经典的拉格朗日对偶理论,可以将原问题变换为对偶问题,这将使得求解最优问题变得更简单. 其对偶问题由如下形式给出:

$$\begin{cases} \max_{\alpha} -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle + \sum_{k=1}^l \alpha_k \\ \text{s.t. } \sum_{i=1}^l \alpha_i y_i = 0, \quad 0 \leq \alpha_i, \quad i=1, \dots, l \end{cases} \quad (3)$$

求解以上优化问题,计算得到

$$w^* = \sum_{i=1}^l \alpha_i y_i x_i, \quad b^* = -\frac{1}{2} \langle w^*, x_r + x_s \rangle \quad (4)$$

式中: x_r 和 x_s 是两类中任意的支持向量. 依据 KKT 互补条件,其中只有少量最靠近超平面样本点的 α_i 值不为零, Vapnik 等人称之为支持向量(SV). 由式(4)可以看出,超平面只是由训练样本集中的一个很小的子集(支持向量集)决定. 最终决策函数为

$$f(x) = \text{sgn} \left(\sum_{i=1}^l \alpha_i^* y_i (\langle x, x_i \rangle + b^*) \right) \quad (5)$$

对于一个给定的特征向量,根据式(5)就可以判定该特征向量对应类别. 对于非线性特征向量,目标函数变为求式(6)方程的最优值.

$$\Phi(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \quad (6)$$

其中惩罚参数 C 作为综合这两个目标的权重. 非线性超平面又称广义最优超平面. 广义超平面的查找方法与线性可分情况几乎完全相同,只是约束条件变为 $0 \leq \alpha_i \leq C$. 最优决策函数的形式与式(5)一样.

2 基于 SVM 的不连续数字语音识别

为了识别语音,需要获取孤立语音信号对应的特征向量. 而要得到特征向量,就需要从一组语音信号中提取孤立的语音信号,即端点检测. 端点检测和特征向量提取是本算法的两个关键步骤.

2.1 语音信号的端点检测

语音信号具有随机和短时平稳性. 不同说话者的语速不同,表达方式也不尽相同,使得相同的语音在不同的环境下有不同的波形与之对应,这就给端点检测带来一定困难. 图 1 是采集的一组语音信号实例(采样频率为 44 100 Hz).

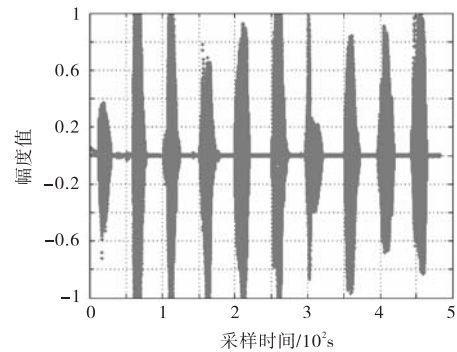


图 1 一组数字语音信号示例

Fig.1 Example of a group of speech signals

为了对图 1 所示语音信号进行端点检测,就需要处理大量数据,这无疑会增加时间消耗. 本文提出的基于短时区域能量谱的方法能够对语音信号进行快速端点检测. 该方法具体描述如下:

语音信号 $y=x(n)$ 的短时区域能量由式(7)求得.

$$E(n) = \sum_{m=-\infty}^{\infty} [x(n)w(n-m)]^2 \quad (7)$$

式中 $w(m)$ 为矩形窗函数.

对于所有采集的数据都使用统一的矩形窗函数进行抽取,得到短时区域能量和抽取位置 n (抽取窗口的中间位置)的关系图,如图 2(a) 所示.

由抽取过程不难得到如下结论:与能量谱中顶点相对应的短时区域能量最大这一结论在原信号图上更为清楚,位于波形的中心附近(误差为矩形宽度的一半)的短时区域,能量最大.

使用相同的方式,继续对短时区域能量谱进行多次抽取,可以有如下结论:采样点数据量逐次减少,而

其能量谱(相应的波峰和波谷的个数)不会发生改变. 图 2(b)是第 5 次抽取结果的图形表示,该图上的波峰和波谷的个数与第一次抽取(图 2(a))完全相同. 为了计算信号在原始采样图中的起始和终止点的位置,只要根据式(7)进行简单的逆运算即可.

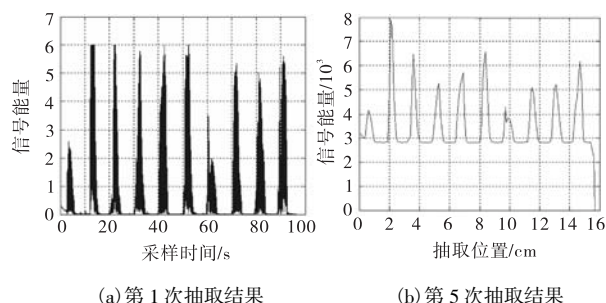


图 2 抽取后的短时区域能量谱

Fig.2 Short-term-region-energy spectrum

2.2 语音信号特征值提取

通过端点检测,能够把单个语音信号从原始信号中提取出来. 然后,要通过对该信号进行特征提取,以便能够找到特征向量,使其能够反映该信号自身特点. 常用的特征提取方法包括:线性预测编码(Linear Prediction Cepstral Coefficients, LPCC)和 Mel 频率倒谱系数(Mel-Frequency Cepstral Coefficients, MFCC);虽然 LPCC 能很好地反映声道特征,但对于语音本身特征却无能为力;MFCC 着重分析人耳的听觉机理,依据听觉特性来分析语音的频谱. 因此,本文选择 MFCC 作为语音信号特征提取的理论基础.

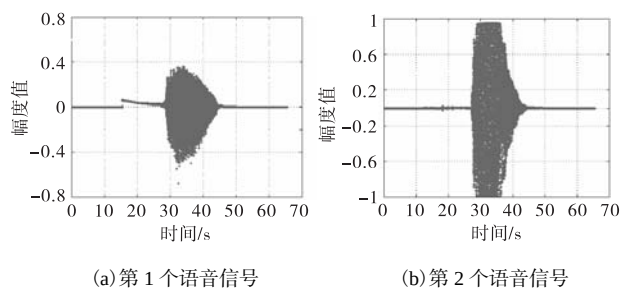
3 实验结果与分析

使用语音处理软件 Cool Edit Pro2.0 对语音进行采集. 采集过程中,要求被采集者以正常的语速说出数字 1—10,每个数字间要有适当的停顿. 然后,将采集到的语音信号保存为*.wav 文件,其中*为该说话者在数据库中的序号.

采样频率为 4 410 Hz,采用立体声模式,16 bit 编码,采集环境为普通实验室,使用加禾-710 M 话筒;采集对象为 60 名男性学生,其中 50 名作为支持向量机的训练者,10 名作为测试者.

3.1 端点检测参数选择及试验结果

端点检测过程中,设置抽样窗口宽度 $w_d=10$,抽样步长 $l_{step}=5$,抽样次数 $n=5$. 每个信号的长度统一为 37 000,不足的部分用零值补充,并且把信号中心放置在窗口中间. 抽取的结果如图 3 所示.



(a) 第 1 个语音信号

(b) 第 2 个语音信号

图 3 端点检测后的孤立数字语音信号

Fig.3 Unitary number speech detected by end-point-checking way

3.2 MFCC 特征参数的提取

将 3.1 抽取后的信号作为输入,在进行抽取特征值之前,还要对该信号进行简单的分帧处理,每一帧信号的长度取值为 256(取 2^n 是为有利于 FFT 等操作). 根据 2.2 所述,把得到单帧对应的特征向量(1×24)进行合并,最后得到的向量作为整个信号对应特征值(帧数 $\times 24$),如图 4 所示.

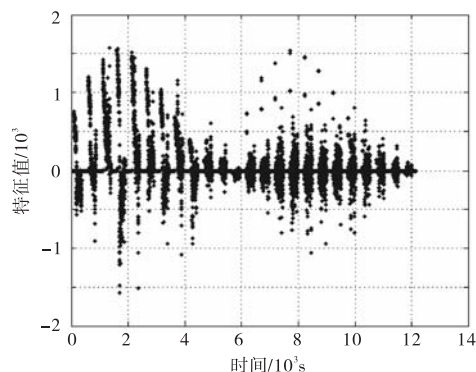


图 4 数字 1 对应的特征向量

Fig.4 Eigenvector corresponding to one number

3.3 最优模型与参数选择

根据 SVM 原理分析,对于非线性样本要采用合适的核函数将非线性特征向量映射到线性空间,然后再使用式(5)符号函数进行分类. 如核函数选择不当,则难以获得满意的效果. 常用的核函数包括径向基核函数(Radial base function, RBF)、多项式核函数(Polynomial)、高斯核函数(Gaussian). 其中,径向基核函数为

$$K(x, x_i) = e^{-\gamma(x-x_i)^2} \quad (8)$$

表 1 为三种核函数对单个语音信号的识别结果. 结果表明, RBF 要优于其他两种核函数.

使用公式(8)径向基核函数,根据实验确定最优参数为:惩罚参数 $C = 1\,000$, $\gamma = 2$. 采用此参数的实验结果如表 2 所示. 结果表明:基于 SVM 的语音识别算法相对于 HMM 语音识别算法速度快,且结果

更准确.

表 1 不同 C-SVM 模型的识别结果

Tab.1 Recognition results of different models of C-SVM

核函数	分析时间/ms	准确率/%
RBF	1.3	99.135
Gaussian	1.5	96.176
Polynomial	1.2	93.152

表 2 采用 MFCC 特征向量和 HMMD 进行识别结果比较

Tab.2 Results of recognition using MFCC eigenvectors and HMMD

序号	准确率/%		分析时间/ms	
	SVM	HMMD	SVM	HMMD
1	99.257	95.5	1.2	5.2
2	99.155	97.3	1.2	7.6
3	99.130	93.2	1.3	6.5
4	99.019	92.3	1.3	8.3
5	98.992	91.4	1.2	7.5
6	98.254	97.5	1.2	6.5
7	100	98.9	1.4	6.4
8	99.8	97.5	1.3	7.4
9	99.7	99.1	1.5	7.5
10	99.3	93.6	1.2	8.2

4 结 语

本文提出并实现了使用短时区域能量谱对语音信号进行端点检测的方法,在此基础上,使用 Mel 频率倒谱系数(MFCC)对检测出的孤立语音信号进行特征抽取,每个语音信号都获取了 $n \times 24$ 个特征向量.在进行语音识别前,先建立了一个由 50 个志愿

者的语音构成的语音库作为训练数据,使用 10 个不同人员的语音作为测试对象.结果表明:基于 SVM 的语音识别算法相对于 HMMD 语音识别算法速度快,且结果更准确.不足之处是对单个语音有一定的应用价值,要求志愿者使用普通话或近似标准话进行采集.在实际的使用过程,由于我国方言众多,没有充分考虑到这一事实,这将是今后的研究重点.

参考文献:

- [1] Davis K H, Biddulph R, S Balashek S. Automatic recognition of spoken Digits[J]. Acoust Soc Am, 1952, 24(6): 634-645.
- [2] Leonard E Baum. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains[J]. The Annals of Mathematical Statistics, 1970, 41(1): 164-171.
- [3] 潘泉, 张磊, 孟晋丽, 等. 小波滤波方法及应用[M]. 北京: 清华大学出版社, 2005: 2-11.
- [4] 唐道南, 秦勇, 沈丽琴, 等. 语音识别技术和应用[J]. 微电脑世界, 1997(11): 41-43.
- [5] Rabiner, L R. A tutorial on hidden markov models and Selected Applications in Speech Recognition[J]. Proceedings of the IEEE, 1989, 77(2): 257-285.
- [6] Cortes C, Vapnik V N. Support vector networks [J]. Machine Learning, 1995, 20: 273-297.
- [7] Burges C J C. A Tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2): 121-167.

(上接第 54 页)

参考文献:

- [1] 樊建明, 陈渊睿. 基于 SHT11 的温室多点测量系统设计[J]. 国外电子测量技术, 2006(11): 4-8.
- [2] 于华丽, 赵晓顺, 刘淑霞, 等. 传感器 SHT71 在温湿环境检测系统中的应用[J]. 农机化研究, 2008(5): 151-153.
- [3] 王永志, 刘媛媛. 大型粮库的温湿度监测报警控制系统[J]. 农机化研究, 2008(8): 167-169.
- [4] 郑长征, 毛哲, 谢兆鸿. 基于 nRF905 的粮库无线测温系统[J]. 微计算机信息, 2007(5): 284-286.
- [5] 辛颖, 谢光忠, 蒋亚东. 基于 Zigbee 协议的温度湿度无线传感器网络[J]. 传感器与微系统, 2006(7): 82-84.
- [6] 郑锋, 王巧芝, 孙西瑞. 温室大棚自动控制系统的设计[J]. 农业科技与信息, 2008(1): 47.
- [7] 孟臣, 李敏. SHT71 数字式温湿度传感器原理与应用[J]. 世界电子元器件, 2003(8): 66-68.
- [8] 杨海东, 陈高平. ISD4004-16M 语音芯片的循环录放电路设计[J]. 单片机与嵌入式系统应用, 2004(12): 54-57.
- [9] 陈慧明, 李艳华, 王静滨. 单片机控制手机的硬软件接口技术及其应用[J]. 微计算机信息, 2005, 21(1): 139-140.
- [10] 李德领, 马潮. 嵌入式系统中短消息实时处理的实现[J]. 单片机与嵌入式系统应用, 2006(1): 33-39.