

用模拟退火算法实现语音识别中的矢量量化¹

王 可 王翠梅*

(四川大学无线电系 成都 610064)

*(西南民族学院计算机科学系 成都 610041)

摘 要 矢量量化在语音识别中有着重要的作用。经典的 K 均值算法收敛速度快,但极易收敛于局部最佳点;其它的一系列改进算法在克服其局部收敛问题的同时,又显著增加了运算量。本文提出了用模拟退火算法实现语音识别中的矢量量化过程,能够较好地协调运算量和收敛质量之间的矛盾。文章讨论了具体算法,并给出了实验数据。结果表明该方法的综合性能优于现有算法,具有较高的实用价值。

关键词 矢量量化, 语音识别, 收敛, 模拟退火算法

中图分类号 TP391.42

1 引 言

在语音识别系统中为了减少识别过程中所需的运算量和存储量,简化识别过程的模型和算法,使之达到实时识别,通常需要对经过预处理得到的观察矢量进行矢量量化^[1-4],产生标量形式的观察序列。

矢量量化方法已有不少^[3,4],包括:K 均值算法(K-Means),分裂法(Binary-Split),K 均值修正算法(MKM),链映射法,共有最邻近法(SSN),迭代自组织数据分析法(ISODATA),预测矢量量化法(PreVQ),乘积矢量量化法以及利用先验知识的 K 均值修正算法和 K 均值的再修正算法(M-MKM)。

其中 K 均值算法是最基本的算法,其运算量最小。缺点在于其初始码本是随机选择的,而初始码本的盲目选择不但可能使收敛缓慢,且会导致码本落入结果不够满意的局部最优解。利用语音先验知识的 K 均值修正算法以及链映射法等都是为了解决此问题而提出的,虽然有助于克服 K 均值算法局部收敛的缺陷,但同时显著增加了计算量,而且对于包含大量样本的训练集合适应性较差。K 均值修正算法和 K 均值的再修正算法为了克服由此所带来的计算量增大的问题,将量化过程中所需的求距离运算一次算好编排成表,在以后的运算中采用查表的方法来减少计算量。这种方法实际是用空间来换取时间,且当总点数较大时,会给距离表的处理带来困难。

模拟退火算法^[5]是一种适合于求解大规模优化问题的技术,特别是当优化问题有很多局部极值而全局极值又很难求出时,模拟退火算法尤其有效。模拟退火算法基于玻尔兹曼概率分布: $P(E) \sim \exp(-E/kT)$ 。即一个处于热平衡状态且具有温度 T 的系统,其能量按照概率分布于所有不同的能量状态 E 之中。即使在很低的温度下,系统也有可能处于一个较高的能量状态,只不过这种可能性很小。参数 k (玻尔兹曼常数)的作用是将温度与能量联系起来。系统的能量可能上升也可能下降,但是温度越低能量显著上升的可能性就越小。米特罗波利斯及其合作者将此原理渗透到数值计算中。他们对于一个模拟热力学系统提供了一系列选择项,并假设系统构形从能量 E_1 变化到能量 E_2 的概率为 $P = \exp[-(E_2 - E_1)/kT]$ 。当 $E_2 < E_1$, 将会出现 $P > 1$, 此时指定概率值 $P = 1$ 。可见该系统的能量总是采取下降过程,但偶尔采取上升步骤。

为了将此算法应用于热力学以外的系统,必须提供以下几项基本要素:(1)对可能的系统构形的一种描述。(2)一个有关构形内部随机变化的生成函数,这些变化作为“选择项”

¹ 1998-06-11 收到, 1999-02-29 定稿

提交给该系统。(3) 一个目标函数 E (类似于能量) 求解 E 的极小值, 即为算法所要完成的工作。(4) 一个控制参数 T (类似于温度) 和一个退火进程, 该进程用来说明系统是如何从高位向低位降低的, 例如在温度 T 时每次下降步骤中要经过多少次随机的构形变化以及该步长是多大等。

本文正是利用模拟退火算法的上述特点, 对矢量量化中的基本算法——K 均值算法加以改造, 建立了模拟退火矢量量化算法 (AVQ)。实验结果表明, 此算法只需较少的运算量及存储量就能实现全局最优点的求解。

2 用模拟退火算法进行矢量量化

矢量量化在训练过程和识别过程中起着不同的作用。语音识别系统中矢量量化的主要工作是聚类, 即从线性预测编码 / 复倒谱 (LPC/Cepstral) 分析^[1,6-8] 所得的特征空间 $T = \{T_1, T_2, \dots, T_N\}$ 中拟定最具代表性的一组点 $R^* = \{r_1^*, r_2^*, \dots, r_M^*\}$, 成为一组聚类中心或码本, 每个中心称为码字。其中 R^* 的生成准则为 $R^* = \min_{\{R\}} \left[\frac{1}{N} \sum_{1 \leq n \leq M} d(T_i, r_m) \right]$; 式中 $\{R\}$ 表示所有可能的码本集, $d(T_i, r_m)$ 为两点之间的距离, 具体定义为两帧语音特征 T_i 和 r_m 之间的失真测度。于是特征空间中的任意一点均可按最小距离准则用码本之一来代表, 从而将特征空间划分成为 M 类。

我们考虑用模拟退火算法进行矢量量化, 将模拟退火中的四个要素定义如下:

(1) 构形: 经 LPC/Cepstral 分析所得的特征空间 $T = \{T_1, T_2, \dots, T_N\}$, 每个元素 T_i 为某一帧语音信号的 p 阶复倒谱矢量 $(C_i(1), C_i(2), \dots, C_i(p))^T$, 一个构形就是特征空间的一个划分。

(2) 调整: 一个矢量从一个类随机转移到另一个类。

(3) 目标函数: 目标函数 E 定义为总的失真测度。两帧语音特征 T_i 与 T_j 之间的失真测度定义为 $d(T_i, T_j) = \sum_{k=1}^p (C_i(k) - C_j(k))^2$ 。因此, $E = \frac{1}{N} \sum_{i=1}^N \min_{1 \leq m \leq M} d(T_i, r_m)$ 。

(4) 退火进程: 这一过程要借助实验确定。首先进行一些随机调整, 利用它们确定调整过程中将会遇到的 ΔE 值的范围, 以及降温的速率。在每次退火进程中, 尝试一定数量的调整步骤, 从中找到有效的转移。当某一温度下无法再形成有效的转移时, 则停止。

利用模拟退火进行矢量量化的框图如图 1 所示, 具体算法步骤如下:

(1) 初始化: 将整个待聚类的矢量点随机划归到 M 个类。

(2) 迭代过程:

(a) 求每类的中心, 将第 i 次求得的第 j 类的中心定义为 C_j^i 。

(b) 收敛性判断: 如果 $\left| \frac{E(k) - E(k-1)}{E(k)} \right| < \alpha$, ($0 < \alpha < 1$) 或者无法继续退火进程, 则聚类收敛, 转向步骤 (3); 否则继续退火过程。其中 $E(k)$ 代表第 k 步得到的总失真测度。

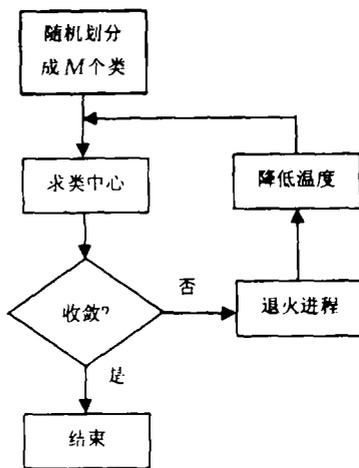


图 1 AVQ 流程

(c) 退火过程: 随机地将第 m 类中的矢量元素 x_k 转移到第 n 类. 该转移有效的条件是: 其转移代价, 即对目标函数的贡献 $dE = -(d(x_k, c_m^i) - d(x_k, c_n^i))$, 满足 $dE < 0$, 或者产生一个 $0 \sim 1$ 之间均匀分布的随机数 a , 满足 $a < \exp(-dE/t)$, 其中 t 表示温度. 重复一定次数的退火步骤, 然后降低温度.

(3) 结果: 最后求得的 M 个中心即为聚类中心, 这些中心形成码本集合.

3 实验结果

本实验采用数字 0 - 9 的语音信号经过 LPC/Cepstral 分析所得的矢量作为原始数据. 每个数字由 4 个人发音 12 次, 共计 120 遍发音. 经过图 2 所示框图进行处理, 得到 2175 个 13 阶复倒谱向量. 具体过程为:

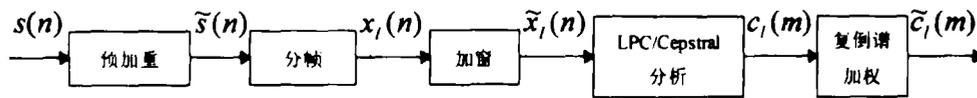


图 2 语音信号的预处理框图

(1) 预加重: 以 10kHz 采样的语音信号 $s(n)$ 经一阶数字滤波器预加重成为 $\tilde{s}(n)$, $\tilde{s}(n) = s(n) - 0.95s(n-1)$.

(2) 分帧: 将语音信号分割成 20ms 的语音帧, 前后相邻帧之间各有 5ms 重叠.

(3) 加窗: 我们采用汉明 (Hamming) 窗来消除由分帧引起的信号边缘锐变. 汉明窗的定义为: $w(n) = 0.54 - 0.46 \cos[2\pi n/(N-1)]$, $(0 \leq n \leq N-1)$.

(4) LPC/Cepstral 分析: 信号经 LPC/Cepstral 分析得到 13 阶复倒谱矢量.

(5) 复倒谱加权: 我们选用正弦窗对复倒谱系数进行加权. $w_c(m) = 1 + (Q/2) \sin(m\pi/Q)$, 其中 Q 为复倒谱阶数.

图 3 给出了对这些向量分别采用 K 均值算法, K 均值修正算法以及模拟退火量化算法进行聚类的收敛情况. 由于 K 均值算法和模拟退火量化算法所产生的结果带有一定的随机性, 因此在表 1 中给出了重复多次的实验结果. 从中可以看出模拟退火量化算法的收敛结果大大优于 K 均值算法, 与 K 均值修正算法接近; 而它所需的计算量又远远低于 K 均值修正算法所需的计算量.

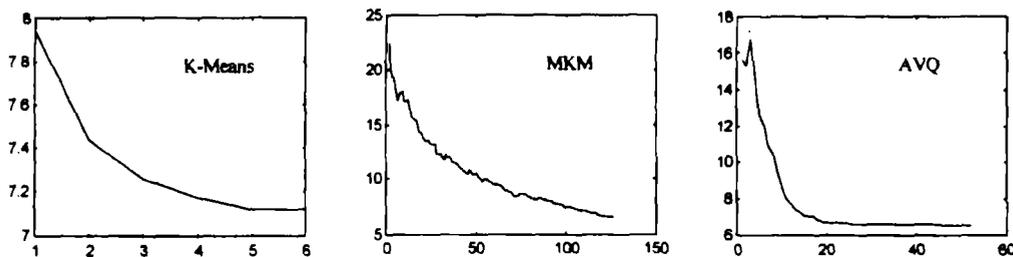


图 3 64 阶矢量量化收敛过程对比 (X 轴: 迭代次数; Y 轴: 总失真测度)

表 1 128 阶矢量量化结果对比

序号	算法					
	K-Means		MKM		AVQ	
	迭代次数	总失真测度	迭代次数	总失真测度	迭代次数	总失真测度
1	7	5.238	213	4.859	63	4.893
2	6	5.559	—	—	59	4.921
3	6	5.244	—	—	58	4.958
4	7	5.129	—	—	59	4.877
5	9	5.262	—	—	61	4.955

4 结 论

本文建立了语音识别系统中矢量量化过程的模拟退火量化算法。该算法具有迭代次数少而且收敛效果较好的特点。同时值得指出的是,本算法的每次调整仅需进行两次距离计算,且整个算法中没有用到其它矢量量化算法中所需的由给定中心进行划分的步骤。因此每步迭代所需的运算量是很小的。实验结果表明该方法的综合性能优于现有算法,具有较高的实用价值。

参 考 文 献

- [1] Rabiner L.R. A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE, 1989, 77(22): 257-285.
- [2] Huo Q. Contextual vector quantization for speech recognition with discrete hidden Markov model. Pattern Recognition, 1995, 28(4): 513-517.
- [3] 胡光锐. 语音处理与识别. 上海: 上海科学技术文献出版社, 1994, 200-297.
- [4] 陈尚勤, 等. 近代语音识别. 成都: 电子科技大学出版社, 1991, 66-85.
- [5] Press W.H. 等著, 傅祖芸, 等译. C 语言数值算法程序大全. 北京: 电子工业出版社, 1995, 378-389.
- [6] Tourneret J. Nonlinearity of a non-linear transformation of AR parameters: Application to reflection and cepstrum coefficients. Signal Processing, 1997, 62(1): 1-14.
- [7] Tohkura Y A. Weighted cepstral distance measure for speech recognition. IEEE Trans. on ASSP, 1987 ASSP-35(10): 1414-1422.
- [8] 马卡尔著, 姜乃英译. 语音信号线性预测. 北京: 中国铁道出版社, 1987, 第 1 章.

ANNEALING VECTOR QUANTIZATION IN SPEECH RECOGNITION

Wang Ke Wang Cuimei*

(Department of Radio Electronics, Sichuan University, Chengdu 610064)

*(Department of Computer Science, Southwest Nationalities Institute of China, Chengdu 610041)

Abstract Vector quantization plays an important role in speech recognition. Traditional K-means algorithm owns the advantage of fast convergence, but it is difficult to get the global optimal result. Some modified algorithms have been proposed to overcome this drawback, but they also increase the computation greatly. In this paper, a new algorithm which is based on annealing algorithm is proposed to compromise the contradiction. In the rest of the paper, the details of the algorithm and related experiments are given. The results demonstrate the algorithm is more effective than other methods.

Key words Vector quantization, Speech recognition, Convergence, Annealing algorithm

王 可: 男, 1972 年生, 硕士生, 主要从事信号处理、计算电磁学和生物电磁学研究。

王翠梅: 女, 1971 年生, 硕士, 主要从事人工智能和模式识别研究。