

## ◎数据库、信号与信息处理◎

# 改进的 SOFM 算法及其在低延迟语音编码中的应用

武淑红<sup>1,2</sup>, 张刚<sup>1</sup>, 张雪英<sup>1</sup>WU Shu-hong<sup>1,2</sup>, ZHANG Gang<sup>1</sup>, ZHANG Xue-ying<sup>1</sup>

1. 太原理工大学 信息工程学院, 太原 030024

2. 太原理工大学 计算机与软件学院, 太原 030024

1. College of Information Engineering, Taiyuan University of Technology, Taiyuan 030024, China

2. College of Computer and Software, Taiyuan University of Technology, Taiyuan 030024, China

E-mail: shh\_wu@163.com

WU Shu-hong, ZHANG Gang, ZHANG Xue-ying. Modified SOFM algorithm and application in low delay speech coding. *Computer Engineering and Applications*, 2009, 45(12): 124-125.

**Abstract:** According to the character of codebook size and codeword dimension in low delay speech coding algorithm, a codebook design algorithm based on modified Self-Organizing Feature Map(SOFM) neural network is proposed. The input train vectors and connection weight vectors are normalized. In order to reduce computation complexity and improve codebook performance, some fast search methods are used in SOFM iterations during searching for the winning neuron and decompose the adaptive adjusting process of network weights into two steps of sequencing and convergence. The proposed algorithm is used to generate vector quantization codebook in low delay speech coding algorithm. Experiment results show that, compared with LBG algorithm, modified SOFM algorithm can greatly improve the synthesized speech quality in the aspect of subjective and objective.

**Key words:** vector quantization; Self-Organizing Feature Map(SOFM) neural network; adaptive adjust; low delay speech coding

**摘要:**根据低延迟语音编码算法训练码书的尺寸和码字维数的特点,提出了一种改进的自组织特征映射(SOFM)神经网络的码书设计方法。对输入训练矢量以及连接权矢量进行归一化,为降低计算量和提高码书训练质量,采用快速的网络学习决定获胜的神经元并对网络权值分阶段进行自适应调整,最后应用于低延迟语音编码中。实验表明,与传统 LBG 算法比较,采用 SOFM 神经网络训练的码书其合成语音的主、客观质量均有较大提高。

**关键词:**矢量量化;自组织特征映射神经网络;自适应调整;低延迟语音编码

DOI:10.3778/j.issn.1002-8331.2009.12.040 文章编号:1002-8331(2009)12-0124-02 文献标识码:A 中图分类号:TP912.3

## 1 引言

对于需要高维大尺寸码书的矢量量化器来说,码书生成过程中的计算量是随着码字矢量的维数  $k$  和码书尺寸  $m$  的增大而急剧增长的,寻求全局最佳码书是十分困难的。实际设计时,必须设法降低复杂度,同时设计高质量的码书来改善矢量量化的性能。

LBG 算法作为矢量量化的基本算法具有经典意义,但由于在训练过程中,总存在少量的离群矢量,使得在训练码书时,码字的分布受到影响,进而使得压缩性能下降。总体来说 LBG 算法存在以下几个缺点<sup>[1]</sup>:一是对初始码书很敏感,初始码书的选择影响码书训练的收敛速度和最终的码书性能;二是在每次迭代的胞腔划分阶段,需要大量存储空间和繁琐的计算;三是码书不能根据训练矢量集数据的随机分布自适应地取得初始码书。

利用自组织特征映射(Self-Organizing Feature Map, SOFM)

神经网络进行矢量量化具有受初始码书影响小、抗信噪误码能力强、适应性强等特点,被广泛应用于图像压缩系统的码书设计<sup>[1]</sup>,但是很少被语音编码系统所采用,特别是未见到有人将其用于低延迟语音编码系统的码书设计。SOFM 是一种具有侧向联想能力的两层结构网络,采用 Kohonen 竞争学习算法,改变获胜节点及相邻节点的权值,实现输入矢量到输出矢量空间的特征映射,同时完成了训练空间的子空间划分,对应的权值的集合即组成码书。

本文结合语音编码的具体算法,针对语音信号的特点提出了两个改善 SOFM 网络性能的方法:一是对输入训练矢量和连接权矢量进行归一化处理;二是将网络权值的自适应调整过程分解成排序和收敛两个阶段,采用不同的学习速率和邻域半径处理各阶段的网络训练。实验结果表明改进的方法使 SOFM 矢量量化器性能得到了明显改善,用于语音编码其主、客观质量

基金项目:国家自然科学基金(the National Natural Science Foundation of China under Grant No.60372058, No.60772101)。

作者简介:武淑红(1969-),女,博士研究生,副教授,主要研究领域为语音编码、计算机应用;张刚(1953-),男,博士,教授,博士生导师,主要研究领域为语音编码与嵌入式系统;张雪英(1964-),女,博士,教授,博士生导师,主要研究领域为语音信号处理。

收稿日期:2008-12-24

修回日期:2009-02-23

比 LBG 算法有较大提高。

## 2 算法原理

### 2.1 基本 SOFM 算法

输入训练矢量集为  $X = \{X_1, X_2, \dots, X_n\}$ , 矢量维数为  $k$ , 码书尺寸为  $m$ 。SOFM 网络有  $k$  个输入节点,  $m$  个输出节点, 对应的连接权矢量  $W = \{W_1, W_2, \dots, W_m\}$  即为码书。输入节点到输出节点的权值为  $w_{ji}, l \in [1, k], j \in [1, m], w_{jl}$  即为码字  $W_j$  的第  $l$  个分量。基本的 SOFM 算法步骤如下<sup>[2-3]</sup>:

(1)  $w_{ji}(0)$  权值初始化: 赋予  $[0, 1]$  区间内的随机值。

(2) 输入训练矢量  $X_i = \{x_{i1}, x_{i2}, \dots, x_{ik}\}, i \in [1, n], X_i$  以并行方式输入到每一个神经元。

(3) 计算输入训练矢量  $X_i$  与输出节点的权矢量  $W_j$  间的失真  $d_{ij}$  作为测度, 选择最小失真的节点  $j^*$  为获胜节点。对节点  $j^*$  及其邻域内的节点的连接权值按下式进行更新:

$$d_j = \min_{1 \leq j \leq m} (d_j) = \min_{1 \leq j \leq m} \left\{ \sum_{l=1}^k [x_{il} - w_{jl}(t)]^2 \right\}, i \in [1, n], j^* \in [1, m] \quad (1)$$

$$w_{jl}(t+1) = w_{jl}(t) + \eta(t) [x_{il} - w_{jl}(t)], l \in [1, k], j \in N_c(t)$$

其中:  $\eta(t) (0 \leq \eta(t) \leq 1)$  为当前迭代的学习速率, 是时间的单调递减函数;  $N_c(t)$  为  $j^*$  的一个欧式邻域, 也是时间的单调递减函数。

(4) 对所有输入训练矢量, 重复步骤(2)至(3)直至算法收敛。

上述算法除确定获胜神经元外, 所有的运算都是并行的。

### 2.2 改进 SOFM 算法

基本 SOFM 学习算法的效率取决于学习速率  $\eta(t)$ 、邻域半径的选择以及连接权值的初始值等, 因此改进算法一般都是通过调整学习速率、邻域半径大小等来提高网络的性能。本文结合语音编码算法的具体情况, 为提高码书训练的质量作了如下改进:

(1) 在基本的 SOFM 算法里, 初始码书是随机产生的, 与原始输入矢量没有任何关系, 不能有效地反映原始数据分布<sup>[4]</sup>, 这里读入前  $m$  个原始数据矢量作为初始码书。

(2) 对输入训练矢量  $X$  和连接权矢量  $W$  用下式进行归一化处理:

$$\bar{s} = s / \|s\| = (s_1, s_2, \dots, s_n) / [(s_1)^2 + (s_2)^2 + \dots + (s_n)^2]^{1/2}, i \in [1, n] \quad (2)$$

在语音编码算法中针对归一化的固定码书重新训练了增益码字使其动态范围与归一化固定码书相匹配。

(3) 学习速率因子  $\eta(t)$  是一个直接影响到 SOFM 算法收敛速度和性能的重要参数。若  $\eta(0)$  为初始学习速率,  $T$  为最大的迭代次数, 当学习速率以  $\eta(t) = \eta(0)(1 - t/T)$  线性变化时, 虽然聚类效果较好, 但是收敛速度慢; 当以指数  $\eta(t) = \eta(0)\exp(-t/T)$  变化时, 聚类效果不是很理想, 但是收敛速度快。对此本文提出了新的调整方式。将式(1)计算的网路权值的自适应调整过程分解为排序和收敛两个阶段:

① 将迭代的前 500 次作为排序阶段, 用来形成权值向量的拓扑排序, 也就是说将初始的无序权值向量排列成有序的二维阵列形式。  $\eta(t)$  初始值接近 1, 且长时间保持在 0.01 以上, 邻域函数的初始半径应包括以获胜神经元为中心的几乎所有神经元, 然后随时间慢慢收缩。由于指数形式的收敛速度快, 从而排序阶段学习速率采用指数形式变化。

② 将迭代 500 次之后作为收敛阶段, 需要微调特征映射以便提供输入空间的准确统计量。学习速率和邻域函数可以按照

基本算法中的线性方式递减。学习速率在相当长的时间内应该保持在 0.01 数量级, 无论如何不允许下降到零, 否则网络会陷入到亚稳定状态。邻域函数在收敛阶段之初应仅包含获胜神经元的最近的邻域, 最终缩减为一个获胜神经元。

### 3 低延迟语音编码算法

低码率、低复杂度的 LD-CELP 语音编码算法在通信领域具有重要的意义, 而现有的编码算法均不满足 ITU 对 8 Kbit/s 语音编码算法的要求<sup>[5]</sup>。这里以降低码率为目的对 G.728 算法进行改进, 提出了一个延迟为 2.5 ms 的语音编码算法。算法以 G.728 结构为基础将激励矢量长度由 5 样点提高到 20 样点, 引入了由最近的历史激励构成的自适应码书和归一化的固定码书的双码书结构, 因此叫作具有自适应码书的 LD-aCELP 算法。

每个输入矢量(20 个样点)构成一个自适应周期(称为帧), 每帧更新一次 LPC 系数。因为采用后向自适应预测技术<sup>[5]</sup>, 向解码器传送的信息只是激励矢量和增益的索引, 因为只有 20 个样点的缓冲延迟, 对于 8 kHz 的采样率就是 2.5 ms 的延迟。

自适应码书搜索时, 先采用后向基音搜索得到基音的初步位置, 然后在此位置的 64 样点邻域内进行自适应码书搜索, 编码占用 6 bit。自适应码书和固定码书增益均用 3 bit 量化, 当固定码书尺寸为 128 和 256 时, 编码算法的码率分别为 7.6 Kbit/s 和 8 Kbit/s。

### 4 实验结果

采用中科院声学所的汉语语音数据文件作为训练数据, 是电话信道采集的含噪数据, 共 1 600 句发声 417 min, 约 200 万个矢量, 其中男、女声各 800 句。分别采用 LBG 算法和改进后的 SOFM 算法训练码书尺寸  $m$  为 128 和 256 的码书, 并以本实验室独立采集的 30 句语音为实验数据, 男、女声各 15 句, 实验数据不包括在训练数据中, 共有 15 489 个矢量进行测试。

分别采用平均分信噪比 SNR 和 ITU P.862 的 PESQ<sup>[6]</sup> 进行测试评价。PESQ 算法得到的评测结果与主观评价指标 MOS 的相关度可以达到平均 0.935, 是现有的基于心理声学模型的客观评价算法中效果最好的指标<sup>[7]</sup>。用标准的 G.728 测试 30 句语音信噪比 SNR 和 PESQ 分别为 18.953 7 和 3.600 5。表 1 为采用 LBG 和 SOFM 码书训练方法得到的码书在低延迟语音编码算法中测试的结果, 从中可以看出, 采用改进的 SOFM 算法训练的码书平均分信噪比 SNR 和平均 PESQ 均比 LBG 算法普遍有所提高, SNR 平均提高 0.97 dB, PESQ 平均提高 0.13。

表 1 不同算法平均分信噪比 SNR 以及平均 PESQ 结果对照表

编码速率/ (Kbit/s)	平均分信噪比 SNR/dB		平均 PESQ	
	SOFM-LD-aCELP	LBG-LD-aCELP	SOFM-LD-aCELP	LBG-LD-aCELP
7.6	17.233 5	16.320 7	3.461 2	3.350 0
8.0	17.843 3	16.824 1	3.558 5	3.413 3

图 1 为各种方案合成语音图, 为一段原始语音信号与采用标准 728、码率为 8 Kbit/s 的 SOFM-LD-aCELP 以及 LBG-LD-aCELP 的不同编码算法相应的合成语音的比较。从图 1 可以看出, 表明在满足 2.5 ms 低延迟的要求下, 编码速率由 G.728 的 16 Kbit/s 降低至 8 Kbit/s, 用改进的 SOFM 矢量量化器编码效果与 16 Kbit/s 的 G.728 十分接近, 明显好于 LBG 矢量量化器。

(下转 156 页)