

基于神经网络的图象编码预测器的实现

汤 剑 徐毓良

(浙江大学计算机系, 杭州 310027)

摘 要 提供了一种基于神经网络的 DPCM 的编码方法。它通过基于图象实例的训练, 可以处理一些线性预测方法所不能处理的高阶特征以及非线性相关性。并由此方法设计了一种典型的预测器—2D-NNADPCM 预测器, 且提出根据不同图象的信息和要求进行人为分块, 以达到更好的编码效果。

关键词 编码, DPCM, 神经网络, 2D-NNADPCM, 分块

1 引 言

在现代生活中, 图象这种信息被大量地应用。但是, 图象的信息量很庞大, 所以储存和传输都非常地困难。因此怎样压缩这些信息, 怎样进行编码就显得非常重要。在编码方法中, DPCM(差值预测编码法)是最常用的方法之一。一般说来, 在用于图象信号的 DPCM 中, 信号通常被认为是一个线性组合, 因此最常用的预测器是线性预测器。而在本文中, 主要讨论一种用于预测编码的神经网络结构。在这种方法中, 信号不再被认为是一线性组合而被认为是非线性组合。

实际上, 预测是对某一时间序列所进行的预测。它根据历史的信息和其它一些相关的信息从而预测出下一个时间点的信息。但是并不是任何一个时间序列都可以进行预测, 它必须符合静态性和遵守线性模型性这 2 个条件。而事实上, 这 2 个条件对于图象序列来说, 并不成立。在本文中, 采取了相应的措施使这 2 个条件成立。对于第 1 点, 先对图象进行分块, 然后对每一个分块进行编码, 虽然并不能完全解决这个问题, 但可以降低一点影响; 对于第 2 点, 由于图象序列具有非线性性, 所以采取了非线性神经网络, 基本上可以解决这个问题。

一个非线性系统的数学模型通常可以用下式表示:

$$g(X_n, X_{n-1}, X_{n-2}, \dots) = V_n \quad (1)$$

其中 V_n 是均值为 0 的白噪声, g 可以是非常复杂的一种非线性表达式。在一般情况下, 该非线性表达式可以展开为泰勒表达式, 如在 $(0, 0, \dots, 0)$ 展开, 则其形式如式(2):

$$X_n = \sum_i a_i V_{n-i} + \sum_i \sum_j \alpha_{ij} V_{n-i} V_{n-j} + \sum_i \sum_j \sum_k \alpha_{ijk} V_{n-i} V_{n-j} V_{n-k} + \dots \quad (2)$$

其中 $\{V_i\}$ 、 $\{V_{ij}\}$ 和 $\{V_{ijk}\}$ 为高斯随机变量, $\{a_i\}$ 、 $\{\alpha_{ij}\}$ 和 $\{\alpha_{ijk}\}$ 为常量系数。原则上, 非线性预测可以通过无限展开来实现。但是实际上, 很难算出表达式中的系数。但是如果应用神经网络的话, 问题就变得简单多了。通过将所需求出的系数作为神经网络的权值或者阈值, 可以很方便地把这些系数求出来。

2 图象模型

假设一幅图象是 $N \times N$ 大小的。如果把图象用 $N \times N$ 大小的矩阵表示, 即每一个象素被一个矩阵元素 X_{mk} 所表示, 则 X_{mk} 表示如下:

$$X_{mk} = \sum_{i=0}^M \sum_{j=0}^K \alpha_{ij} X_{m-i, k-j} + V_{mk} \quad (3)$$

其中, 若 $i=j$, 则 $i=j \neq 0$, V_{mk} 为白噪声。这个式子也正是预测编码的理论前提, 它表示, 当前的一个象素值可以根据先前象素的值进行估计。在该图象模型中, 当前的象素值实际上是先前的象素值和白噪声

的线性组合,因此它要求信号序列必须满足线性性这个条件。但是事实上,该条件是不成立的。为了改善这个不足,我们采用了非线性的模型。其数学表达为:

$$X_n = \sum_i a_i X_{n-i} + \sum_i \sum_j a_{ij} X_{n-i} X_{n-j} + \sum_i \sum_j \sum_k a_{ijk} X_{n-i} X_{n-j} X_{n-k} + \dots + V_n \quad (4)$$

其中, $\{a_i\}$ 、 $\{a_{ij}\}$ 和 $\{a_{ijk}\}$ 为常系数(对于某一短时间而言), V_n 为白噪声。

虽然从理论的角度上来说,如果要获得非常好的预测效果的话,只需要增加尽可能多的高阶项即可,但是在实际运用的过程中,如果无限地增加高阶特性的话,将引起数据爆炸。所以在实际应用中,一般都只取到三阶的特性。

3 神经网络的预测器结构和设计

运用传统算法来寻找一个非线性预测器是相当困难的,但是神经网络的出现给我们提供了一个新思路。我们可以通过设计一个合适的神经网络结构和采用对实际数据的训练来完成非线性预测器的设计。如果训练过程能够实现一个能使预测剩余能量最小的下降梯度的话,那么一个最优预测器也就实现了。

在神经网络结构的选用上,我们采用带有一个隐含层的BP网络。该网络的典型结构如图1所示。

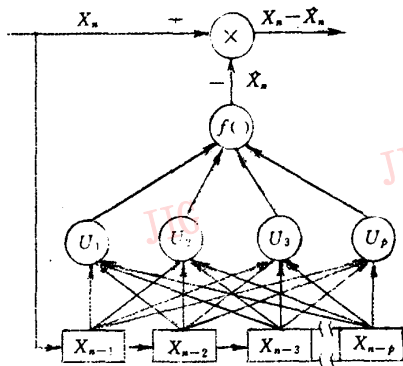


图1 带一隐含层的BP网络结构图

Fig.1 Architecture of BP Neural Network with a hidden layer

在实际应用中,预测值可以由下式表示:

$$\hat{X}_n = \sum_i w_i X_{n-i} + \sum_i \sum_j w_{ij} X_{n-i} X_{n-j} + \sum_i \sum_j \sum_k w_{ijk} X_{n-i} X_{n-j} X_{n-k} + \dots \quad (5)$$

其中, \hat{X}_n 为当前象素的预测值, $\{w_i\}$ 、 $\{w_{ij}\}$ 和 $\{w_{ijk}\}$ 为神经网络中的权值,该神经网络预测器的结构图如图2所示。

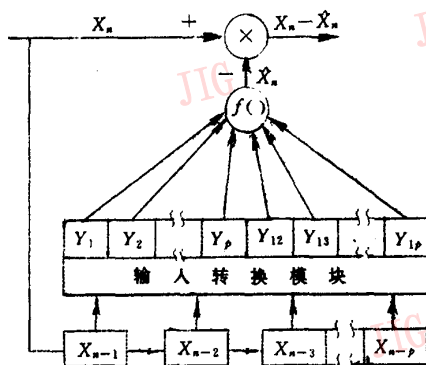


图2 神经网络预测器结构图

Fig.2 Architecture of Neural Network Predictor

从图中可以看到:每一个预测值的计算与先前 P 个值有关,这 P 个值在输入转换模块中进行非线性的组合,而后再与权值相乘形成式(5)所表达的结构。神经网络的权值就相当于传统图象模型中的常量系数 $\{a_i\}$ 、 $\{a_{ij}\}$ 和 $\{a_{ijk}\}$,它表述了神经网络的特征,它是基于实际图象的训练而被最后确定的。因此,神经网络预测器的优化与否的问题就转换为怎样求出这些权值而使预测剩余能量最小化。

预测剩余能量可以用下式来表示:

$$RE = \sum_n e_n^2 = \sum_n (X_n - \hat{X}_n)^2 \quad (6)$$

其中, n 是图象中的所有的点。

神经网络的学习规则有很多种,在此我们采用最常见的梯度下降法。在具体的学习过程中,我们可以让学习过程一直重复至权值的变化值小于某一定值或者预测剩余能量小于某一定值。这2个定值的大小可以根据不同的情况选取。

4 各向异性的处理方法

实际上,上述的二维数学模型,最适合于各向同性的图象,或者更严格地说,适合于在水平和垂直方向上有相同相关特性的图象。然而对于很多图象来说,水平方向和垂直方向的特性并不相同。因此若采用这种方法,对于某些特殊图象,有可能会造成较大的图象偏差。

解决这个问题的方法有2种。第1种方法就是在取已有象素值作为输入时,垂直方向和水平方向

取点个数针对具体的图象可以有所不同。如象素值变化较快或者较没有规律的方向上可以适当地少取一些。第2种方法就是根据图象象素值变化的快慢、规律性、相关性等等,进行人为的分块,从而再对每1块进行编码。这2种方法都带来了相应的困难或麻烦。第1种方法需要不断地改变神经网络的结构,而第2种方法则需要人为地参与。但是相比较而言,第2种方法则显得更为可行一点。

通常在实际应用中,我们均采用固定的分块大小(如 $32 \times 32, 64 \times 64$ 等)来进行编码。但是有些有特殊要求的图象的编码可以采用上述的第2种方法,这样的话,重建后的图象有可能同时获得图象峰值信噪比的增大以及编码率的减少。比如某一图象,背景大体不变,且占据一定的比例,但是又要求该图象中物体细节在重建中不能有大的偏差。这时候采用这种方法就显得有一定的优点。在具体处理的时候,对于背景,则采用较大的图块进行编码,而对于细节,则采用较小的图块进行编码。这样,往往会获得意想不到的效果。

5 2D-NNADPCM 及其仿真结果

2D-NNADPCM(二维神经网络自适应差分预测

测编码法)就是将上述的神经网络预测器运用到局部的图象块中而实现编码的方法。它是一种二维的预测,图3为预测的选取点示意图。

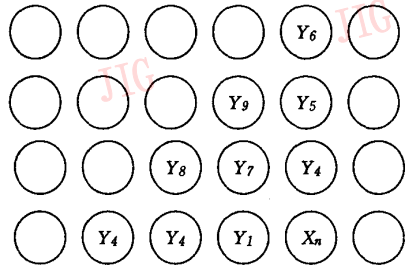


图3 预测编码所选取的先前点示意图
Fig. 3 Geometry of Prediction schemes

对于每一图块,神经网络预测器被自适应地训练,从而形成一组权值。一般情况下,训练过程在3~4个回合后结束。考虑到各分块之间的相关性,上图块训练后的权值被用于下一图块的训练时权值的初始值,这样可以节省时间。

在该预测方法中,预测差值被量化为3个等级,而后再对其进行长度为5的霍夫曼编码。我们采用如图4所示的用于2D-NNADPCM的神经网络结构作为预测器结构。

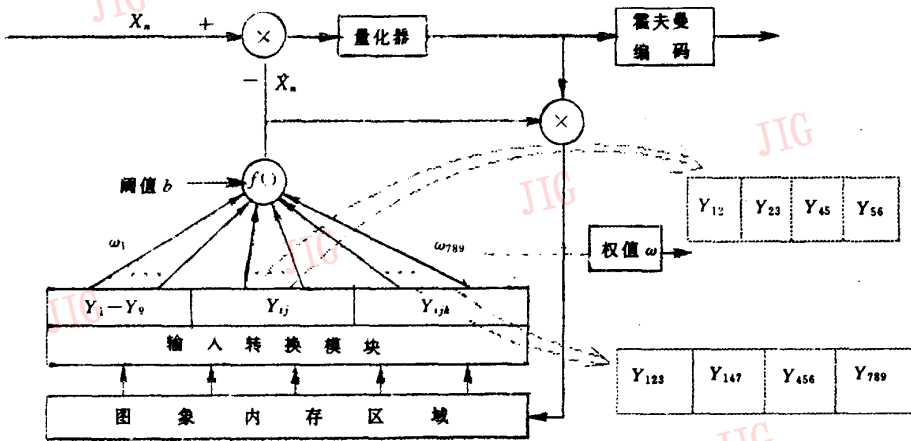


图4 实际应用中2D-NNADPCM的神经网络结构
Fig. 4 Architecture of Practical 2D-NNADPCM Neural Network

其中,

$$\hat{X}_n = f(s + b) = \frac{1}{1 + e^{-(s+b)}}$$

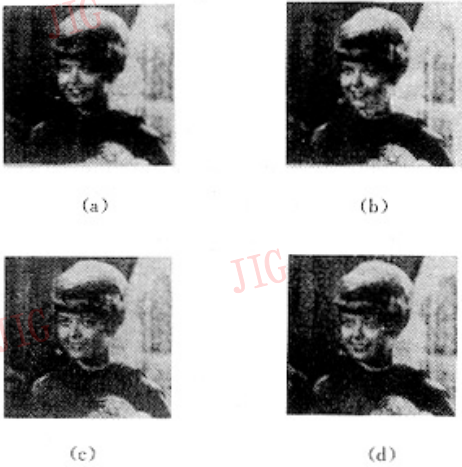
$$s = \sum_{i=1}^9 w_i Y_i + w_{123} Y_{123} + w_{147} Y_{147} + w_{456} Y_{456} + w_{56} Y_{56} + w_{789} Y_{789}$$

$$Y_{12} = Y_1Y_2, Y_{34} = Y_3Y_4, Y_{45} = Y_4Y_5, Y_{56} = Y_5Y_6,$$

$$Y_{123} = Y_1Y_2Y_3, Y_{147} = Y_1Y_4Y_7, Y_{456} = Y_4Y_5Y_6, Y_{789} = Y_7Y_8Y_9$$

由于大的分块容易造成图象的失真,而过小的分块又会造成所需权值等的增加而带来不必要的消耗。所以在对图 5(a)所示的“女郎”图象(大小为 256 × 256)的仿真实验中,我们采用的固定分块大小为 64 × 64。实验结果显示平均编码率为 0.41 bit/pel。另外在传输的过程中,还需要传输权值和阈值,因此必须对它们也要进行编码。对于每一个分块,一共有 17 个权值和一个阈值共 18 个系数,而每一个系数需要 5 个 bit,因此总的编码率为 0.435 bit/pel。

图 5(c)为用该方法重建后的图象,而图 5(b)为用线性 DPCM 方法编码(1bit/pel)的重建图象,使用该编码方法后的重建图象比使用线性 DPCM 后的重建图象要有明显的改善,而且没有变形。



(a)原图 (b)线性 DPCM 编码后的重建图(1 bit/pel) (c) 2D-NNADPCM 编码后(d)带人为分块的 2D-NNADPCM 的重建图象(0.435bit/pel)编码后的重建图象(0.41bit/pel)

(a) Original Image (b) Reconstructed Image with Linear DPCM(1 bit/pel) (c) Reconstructed Image with 2D-NNADPCM (0.435bit/pel) (d) Reconstructed Image with 2D-NNADPCM with sub-block(0.41bit/pel)

图 5
Fig. 5

另一个衡量编码性能的指标是峰值信噪比 SNRP,它定义如下:

$$SNRP = 10 \log \frac{(225)^2}{\sigma_q^2} \text{dB} \quad (7)$$

其中, σ_q^2 是原始图象和重建图象象素的平方差。

用该 2D-NNADPCM 方法处理后的峰值信噪比大约为 25.8dB。

此外,考虑到各向异性,我们按图 6 所示的方式对该图象进行分块。并且按分块对其采用 2D-NNADPCM 进行编码。其中有 3 个区域为非 64 × 64 大小:

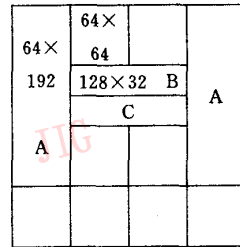


图 6 “女郎”图象的人为分块示意图
Fig. 6 Sub-Block of the Image "Girl"

区域 A——背景部,图象象素值大体相同,且图象包含的信息量不重要,故把原来的 3 块合为 1 块;

区域 B——额部头发等,纵向相关性,故划分成 1 块进行单独处理;

区域 C——脸颊部,包括眼和鼻,属细节要求高部分,故也进行单独处理。

分块后再进行 2D-NNADPCM 编码的仿真实验结果显示该编码方法的平均编码率为 0.39bit/pel。再加上系数以及分区信息的码值,总的编码率为 0.41 bit/pel。图 5(d)为用该方法编码后的重建图象,该图象比图 5(b)和图 5(c)都有所改善。用该方法处理后的峰值信噪比为 32.2dB,也有一定的改善。

6 结束语

本文中所提到的基于神经网络网络结构的预测器是一个较为新颖且很有效的预测器。它可以处理一些线性预测器所不能处理的高阶特征以及非线性相关性。实际上仿真实验的实验结果如:平均编码率、峰值信噪比和重建图象也表明了该方法的优越性。另外,实验同时还表明人为分块后将进一步提高各项指标。

参考文献

1 Rosenfeld Kak. Digital Picture Processing, 1976.

2 Priestley M B. Non-linear and non-stationary time series analysis, Academic Press, 1989.

3 姚庆栋等. 图象编码基础, 杭州, 浙江大学出版社, 1993.



汤剑, 1974年生, 1991年进入浙江大学电机工程系本科学习, 1995年进入浙江大学计算机科学与工程系应用专业进行硕士研究生学习。主要从事图象, 图形, GIS等方向的研究。



徐毓良, 男, 61岁, 教授, 浙江大学人工智能研究所工作, 从事智能控制和图形仿真研究。

An Image Encoding Predictor Based on Neural Network

Tang Jian, Xu YuLiang

(Department of Computer, Zhejiang University, HangZhou 310027)

Abstract In this paper, a neural network approach to DPCM (differential pulse code modulation) design for encoding of images is presented. This approach can deal with high-order statistics and nonlinear correlation which the linear method can not . Based on this method, this paper provides a predictor —2D-NNADPCM, and presents an approach that divides an image into different sub-block according to different messages and requirements in order to achieve better results.

Keywords Coding , DPCM, Neural network, 2D-NNADPCM , Sub-block

《OpenGL 编程指南》

孙绍麟 费月娥 编译

全书共十二章,三个附录,16开,300页。全面、深入、详细地讨论了OpenGL编程中的实际问题,对OpenGL编程人员极具参考价值。

内容包括:绪论,绘几何对象,坐标变换,显示表,颜色,光照,混合、反走样和雾,绘象素、位图、字体和图象,纹理映射,帧缓存,鉴别器和NURBS,选择和反馈,附录1,2,3。定价:48元

邮购:《中国图象图形学报》读者服务部(100088,北京海淀区花园路6号,

电话:62378784,联系人:李如珍)