

神经网络在图像压缩技术中的应用^{*}

刘春阳 梁德群 宋焕生 吴更石

(西安交通大学电信学院, 西安 710049)

摘 要

综述了神经网络作为图象压缩信号处理工具的方法。由于神经网络的大规模并行及其分布结构, 使之成为解决数据压缩的有力工具; 网络的特征与人类视觉系统的特征相类似, 这就使我们更容易处理视觉信息。例如, 多层感知可作为差分脉冲编码调制的非线性预测器, 已证明这种预测器较线性预测器可改进预测效果。另一活跃的研究领域是用 Hebbian 学习规则获取主分量, 主分量是理想的线性 KL 变换的基向量。这些学习算法的计算更优越于传统特征分解技术并适应于输入信号的变化。还有另一种模型为 SOFM (the Self-Organizing Feature Map) 已在向量量化码本设计中有许多成功的应用。结果码本与标准 LBG 算法相比对初始条件的依赖程度较低, 且码字的拓扑序可用来提高编码的效率并减低计算的复杂性。

关键词 神经网络, 图像压缩, 自组织特征映射

分类号 AMS(1991) 94A 13; CCL O 236

1 引 言

随着信息技术, 特别是通信技术的发展, 图象压缩的重要性日益突出, 传统的方法无法满足需要。许多学者正在探索新的更有效的方法, 其中用神经网络进行图象压缩是最为活跃的方向。从计算方式的角度看, 人工信息处理系统可分为二类: (1) 序列计算方式, (2) 并行处理方式。人工神经网络的研究表明特别适应于并行处理, 而在哺乳动物的视觉系统中, 视觉信息是以大量平行的内部相关的网络来处理。这种平行结构即从视网膜到视皮层的高度有序的结构是很明显的, 正是人工神经网络可以发挥其优点的地方。

近来人们将神经网络用于图象压缩取得较好的效果。对于预测编码, 变换编码, 向量量化等主要三种编码方法使用人工神经网络比序列计算方式显示了很大的优越性。这主要表现在: (1) 由于网络的权是由训练产生的, 可以通过在处理新数据过程中继续训练, 使它适应输入数据的变化。(2) 数据是单独地被训练, 就不需要过量储存全部训练集, 这在处理极大量的数据如图像时尤其重要。(3) 由于人工神经网络的高度连结可使神经网络自我组织。(4)

^{*} 本文 1995 年 12 月 27 日收到。

神经网络与神经生物系统之间的相似性使人工神经网络更接近人类视觉信息的处理方式。

为能更清楚的简述人工神经网络对这三种方法的改造效果, 本文将简述如下问题: 在第二部分中主要综述预测编码及应用神经网络的预测编码方法, 即应用神经网络作为非线性预测器。第三部分讨论变换编码及应用神经网络的变换编码方法, 这些方法包括线性原理成份分析(PCA)方法, 它应用 Hebbian 学习规则的自相关编码和自适应编码。第四部分讨论矢量量化及自组织特征映射(SOFM)的应用和它的矢量量化的变化。图象压缩技术的综合评价将在第五部分讨论并结束本文。

2 应用神经网络进行预测编码

1) 预测编码 该方法是利用图象在相邻样本之间呈现出高度相关性的特点, 减少其冗余度以达到压缩目的, 可以通过应用预测编码或多层次脉冲编码调节(DPCM)来完成。

图1显示这种系统的框图, 预测器应用历史样本 $x(n-1), \dots, x(n-p)$ 或者图象相邻像素样本, 来测算现有(新)样本的估计: $\hat{x}(n)$ 。正确值 $x(n)$ 与估计值 $\hat{x}(n)$ 之间的差, 即 $e(n) = x(n) - \hat{x}(n)$ 被用来贮存或传送, 随着预测器的精度的提高, 差 $e(n)$ 减小, 导致一个较高的预测结果和一个较高的压缩比。

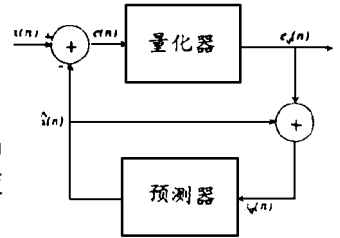


图1 DPCM 框图

最常用的为应用统计模型导出的自回归模型, 如 P -阶自回归模型即:

$$x(n) = \sum_{j=1}^P w_j * x(n-j) \quad (1)$$

公式中 $\{w_j\}$ 应满足:

$$RW = d \quad (2)$$

其中 $[R]_{ij} = E[x(j) * x(i)]$ 是自相关协方差矩阵的第 i 行, 第 j 列, 而 $d_j = E[x(n) * x(j)]$ 为协方差向量 d 的第 j 个元素, 知道 R 和 d , 可以计算未知的系数向量 w 。

2) 多层感知器 当设计一个非线性预测器时, 其目的就是对于给定的前 P 个输入的非线性函数找出理想的参数集 w , 如(3)式所示。

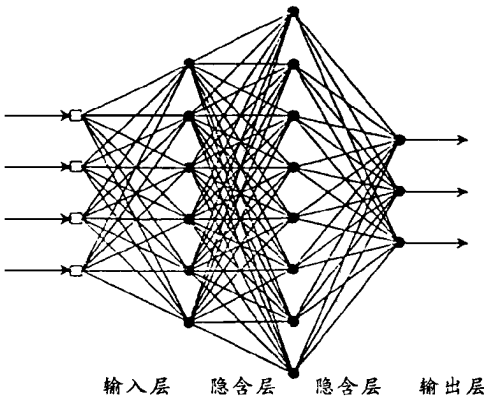


图2 多层感知器

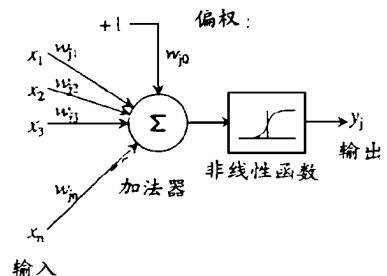


图3 一个神经元模型

$$\hat{x}(n) = f(x(n-1), \dots, x(n-p), W) \quad (3)$$

使这时预测差的均方值 $E[(x - \hat{x})^2]$ 达到最小。如图2所示, 应用多层感知器可以估算出这一类非线性函数。基本运算单元——“神经元”, 它是由“突触”权集, 每个代表一个输入, 加

上一个偏权值数, 一个加法器以及非线性函数组成, 此非线性函数指的是活性函数, 如图 3 所示。每一运算单元计算输入的加权和加上偏数 θ , 然后代入活性函数计算出输出值, 如(4)式:

$$y_j = f(\sum w_{ji} * X_i + \theta) \quad (4)$$

这里 X_i 是此神经元的第 i 个输入值, w_{ji} 是相应“突触”的权。“活性”函数 $f(\cdot)$ 映射变化无限范围的加权和到有限值(有限范围), 一般使用的活性函数是由逻辑函数定义的 S 型函数, 即

$$f(V) = \frac{1}{1 + e^{-V}} \quad (5)$$

如图 4 示。在多层结构中, 计算单元的某一层的输出形成下一层的输入。第一层的输入看作是网络的输入, 而最后层的输出为网络的输出。

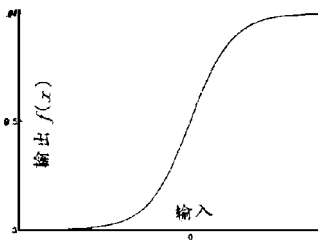


图 4 非线性函数

应用反向传播算法训练网络常可以算出网络的权。反向传播算法是一种监督学习算法, 它能使在方差能量面上用梯度下降法达到最小值。在多层感知器上, 应用这种方法的关键是通过将误差经网络反向传播计算隐层节点的误差值, 在输出层上, 第 i 神经元的局部梯度, $\delta_i(P)$ 如下公式计算

$$\delta_j = y_j * (1 - y_j) * (d_j - y_j) \quad (6)$$

这里 y_j 是第 j 元输出, 而 d_j 是期望响应。对隐层元而言, 第 j 神经元的局部梯度计算如下

$$\delta_j = y_j * (1 - y_j) * \sum_k \delta_k * W_{jk} \quad (7)$$

这里对固定的 j, k 取所有下层的元。一旦算出局部梯度, 就可根据按(8)式修正每一个权值。

$$W_{ji}(t+1) = W_{ji}(t) + \eta * \delta_j(t) * y_i(t) \quad (8)$$

这里: η 是学习率参数, t 是时间, 通常用(8)式的修正合并动量项, 并帮助加速学习过程 [1, 5]。

$$W_{ji}(t+1) = W_{ji}(t) + \eta * \delta_j(t) * y_i(t) + \alpha * [W_{ji}(t) - W_{ji}(t-1)] \quad (9)$$

这里 α 是动量项其取值范围为 $0 < \alpha < 1$ 。

多层感知器可用作非线性预测器, 输入由前面的 p 个数据值构成, 输出为单一个预测值, 在输入与输出层之间是许多不同大小的隐含层, 此如图 5 所示。

由于网络的非线性特征, 神经网络预测差的方差将比线性预测的要小, 这使 DPCM 系统的预测性能增加。Dianat [6] 等使用的网络是: 3 个单元的输入层, 含 30 个单元的隐层和一个输出单元。他们的实验表明, 网络与应用相同输入结构的理想线性预测编码系统相比其信噪比改进 4.1 dB, 而且他们还证明差商从 4.7 改进为 3.9 bpp。

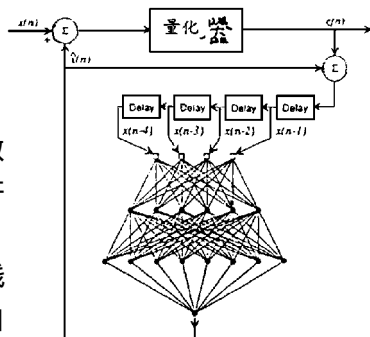


图 5 基于多层感知网络的 DPCM

3) 高阶预测器 另一种方法就是利用后传播算法的梯度下降特点来计算一个特定的非线性模型的优化预测器。Mauikopoulos [7, 8] 已经把基于广义的非线性预测模型的离散时间 Volterra 展开使用到一种非线性预测器上。使用非线性方法的理由是线性 AR 图象模型

没有充分考虑到界线结构如图象边缘,因此对于广义回归模型就需要高阶项。

$$x(n) = \sum_i v_i * x(n-i) + \sum_{i,j} v_{ij} * x(n-i) * x(n-j) + \sum_{i,j,k} v_{ijk} * x(n-i) * x(n-j) * x(n-k) + \dots + \epsilon_i \quad (10)$$

这里 $\{\epsilon_i\}$ 是随机变化的零均值数列。

单层神经网络用作非线性预测器,输入由前 p 样本加上所有的高阶交叉项组成。图6示4阶($p=4$)系统的结构。通过对两幅图象测试表明,用4点相邻的一维预测器在SNR方面与理想线性预测器比较分别改进了4.17 dB和3.74 dB。用9点相邻的二维预测器,在0.51 bpp情况下取得SNR为29.5 dB。

与图6相似的网络是由Pao[9]在1989年提出的,他引入了函数连接网络。此方法用有高阶项的单层网络来代替多层感知器,已证明单层高阶网络与等价的多层感知器相比其学习速度有显著的提高。

4) 串联性再生网络 通过使用管线再生神经网络(PRNN),可将神经网络预测器推到非线性和非静止的信号模型中。PRNN网络由许多微型组件组成。其中每一个组件是带有 $N-1$ 个反馈输入的再生神经网络。图7表示这一系统的结构。每一组件将输入信号的 p 延滞式和 $N-1$ 反馈输入作为其输入。每一组件的输出作为下一层的输入。以这种方式,网络实现了输入的一套非线性函数。最终的网路输出是组件输出的线性加权和。用梯度下降算法训练每一组件的权 w ,又通过标准最小均方差算法[11]计算线性的和的权。在语言信号实验中Haykin和Li[10]发现基于PRNN的预测器比线性FIR滤波器能提高3 dB的预测增益。应用PRNN压缩图象还在探索中。

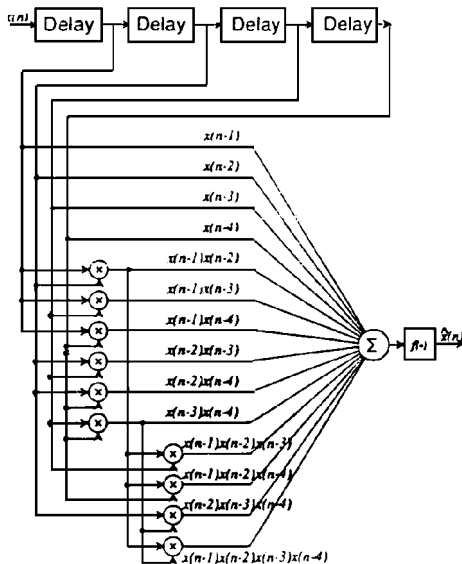


图6 基于4阶非线性AR模型

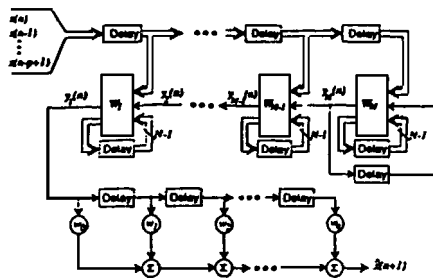


图7 “管线”再生神经网络

3 使用神经网络进行变换编码

1) 变换编码 图象压缩的另一个方法是应用变换使图象产生系数集,选择这些系数的子集并量化以便通过信道传送或贮存,这个方法的目的是选择一个变换,它的一个系数的子集足可满足重建一幅有最小失真的图象。

简单且有效的一类变换编码技术是线性块变换编码,一幅图象分隔成不重叠的 $n * n$ 象素区块,它可看作是 N 维向量 X 这里 $N = n * n_0$ 。线性变换可以写成 $M * N$ 维矩阵 W , $M \leq N$,它在每一块用 W 的 M 行操作。 w_i 看作是变换的基础向量, M 维系数向量 Y 可用公式(4)计算。

$$Y = W * X$$

(11)

如果基础向量 w_i 是正交的即满足(12) 式

$$w_j^T * w_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (12)$$

那么, 逆变换矩阵就是变换矩阵的转置。重构向量为:

$$\hat{x} = W^T * Y \quad (13)$$

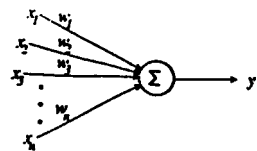
考虑减低均方差(MSE), 理想的线性变换是KL 变换, 变换矩阵 W 由 M 行特征向量组成, 它对应着样本自相关协方差矩阵的 M 个最大特征值, 即 $\Sigma = E[xx^T]$ 。KL 变换与此方法类似产生非相关系数, 并因此而产生最有效的数据编码, 这是因为相邻向量之间的高度相关性的冗余度被去除了。又因为基础向量也是数据的主分量, 所以 KL 变换与主分量分析(PCA) 有关, KL 变换是正交变换, 它的逆变换只是它的转置。当试图实现上述算法时仍存在许多实际困难, 因为图象的自相关协方差矩阵的计算是庞大的, 并需要大量的存贮。另外, 特征值与特征向量的计算复杂度较高, 正变换与逆变换的计算在每一个图象块的量级是 $O(M * N)$, 由于这些困难, 固定基变换可专门用于实现区块变换的方案, 如余弦变换(DCT) 它的计算量级为 $O(N * \log N)$, JPEG 已经采用了线性区块变换编码方法, 这是因为它用标准的DCT 作为变换。

2) 线性 PCA (主分量分析法) 应用神经网络模型的迭代技术是解决通过协方差估计特征分解计算基向量的一种方法。这种方法仅须很少的比特数而计算却很有效。此外, 此方法在图象统计中能适应长期变化。

在1949年, Donald 和Hebb 提出[12] 在神经生物网络中, 修正相联的神经元间的突触强度可影响学习过程, Hebb 的学习假设阐明: 当一个神经元不断参加于激活其它神经元, 则这一神经元引起另一神经元“激活”的能力就增加了。换句话讲, 当输入和输出的神经元同时“激活”, 它们之间的连接就被加强了。

对于人工神经网络, 如图8所示, 神经元之间的相互作用可模仿成简化的线性计算单元。神经元的输出 y 是突触权 $\{w_1, \dots, w_n\}$ 对输入 $\{x_1, \dots, x_n\}$ 的加权和, 或以向量表示为:

$$y = w^T * x \quad (14)$$



用输入和输出值表示“激活率”, 这种模型应用Hebb 学习假设将意味着 x_i 和 y 的值相关时, 权 w_i 将加强。将这一原则推广到同时包含负值的情况时(类似生物网络的抑制作用), 将按输入向量 x 和输出 y 之间的相关性来修正权 w 。

简单的Hebb 规则是相应于输入和输出值的积来改进权

$$w(t+1) = w(t) + \alpha * y(t) * x(t) \quad (15)$$

这里 α 是学习率参数。然而这样的学习规则是不稳定的。这是因为权值将没有边界的增加。在每一步对权归一化可加强其稳定性。如

$$w(t+1) = \frac{w(t) + \alpha * y(t) * x(t)}{w(t) + \alpha * y(t) * x(t)} \quad (16)$$

这里 $*$ 表示欧氏模(范数)。已证明这一规则收敛于输入 x 的最大主分量。Oja 使用级数展开将(16) 式线性化, 成为:

$$w(t+1) = w(t) + \alpha * [y(t) * x(t) - y^2(t) * w] \quad (17)$$

方程(17) 被证明收敛于最大主分量[5]。

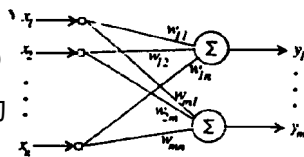


图9 M 主分量线性网络

3) 广义Hebb算法 Oja法则是Hebb学习规则扩展到同时求第一 M 主分量的三基础, 此系统结构如图9所示。每一个输出 y_i 对应着第 i 个主分量神经元的输出。用向量表示为

$$y = W * x \quad y \in \mathbf{R}^M \quad W \in \mathbf{R}^{M \times N} \quad M \leq N \quad (18)$$

Sanger的GHA (Generalized Hebbian Algorithm) 将Oja模型推广到计算 M 主分量, 因为任何主分量的计算均与修正的第一分量的计算相同, 即用Gram-Schmidt正交化去除此主分量与前面主分量的相关性来修正数据。换言之, 第 m 主分量是 w_m 的第一主分量, 这里

$$w_m = w - W_{m-1}^T * W_{m-1} * w \quad (19)$$

W_{m-1} 是 $(m-1) \times N$ 矩阵, 它的 $m-1$ 行是由前 $m-1$ 主分量组成, 将正交性应用到学习法则中形成:

$$W(t+1) = W(t) + \alpha(t) * (y(t) * x^T(t) - L T[y(t) * y^T(t)] * W(t) \quad (20)$$

这里 $L T[\cdot]$ 是下三角算子, 即它的所有对角线以上的元素均为零。在条件 $\lim_t \alpha(t) = 0$,

$\sum_{t=0} \alpha(t) = \infty$ 条件下, W 收敛于其行是 M 主分量的矩阵[13]。

为了评估性能, Sanger使用 8×8 输入块和8维数组输出来实现算法。用非重叠图块扫描两次图象在 512×512 图象上训练网络。学习参数 α 取值范围为 $[0.1, 0.01]$ 之间的给定数。系数是采用非均匀量化。比特分配随每个参数方差变化而变化。在压缩比为0.36 bpp时, 得出标准均方差为0.034。当用同样的 W 矩阵对第二幅独立的图象进行编码时, 在压缩比为0.55 bpp时标准均方差为0.023。Sanger还将此方法用于特征分割及感知模型领域。

4) 自适应主分量提取 Sanger的方法仅使用前反馈连接来计算 M 主分量。Foldiak提出的另一种方法是使用“anti-Hebbian”反馈连接来排除分量的相关性, 此方法的合理性是基于Barlow和Foldiak在视皮层方面的早期研究。在这个方法以及Oja的研究基础上Kung和Diamantaras发展了级数解称为自适应主分量提取(APEX), 其中第 m 主分量的输出 y_m 能由前 $m-1$ 分量以(21)式和(22)式计算得出。

$$Y = W * x \quad (21)$$

和

$$y_m = w^T * x + c^T * Y \quad (22)$$

这里 y 是前 $m-1$ 分量的向量, W 是前 $m-1$ 分量的权矩阵, w 是第 m 分量权向量, c 相当于“anti-Hebbian”中从第 m 分量中去掉前 $m-1$ 分量, 此结构如图10所示。学习法则为:

$$\Delta w = \alpha * (y_m - y_m^2 * w) \quad (23)$$

$$\text{和} \quad \Delta c = -\beta * (y_m * Y - y_m^2 * c) \quad (24)$$

Kung和Diamantaras已经证明权 w 收敛于第 m 主分量, 并算出前 $m-1$ 分量。当网络收敛时, “anti-Hebbian”权 $c(t)$ 收敛于零。理想学习参数 $\alpha \beta$ 可用(25)式计算。

$$\alpha = \beta = \left(\sum_{i=1}^n y_i^2 \right)^{-1} \quad (25)$$

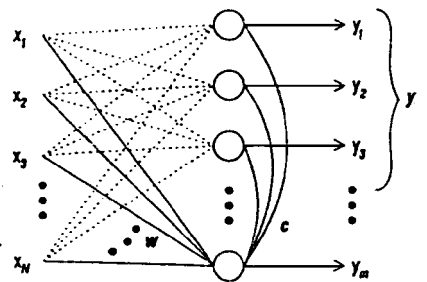


图10 APEX算法网络

这里 n 为输入模式数。如此选择的学习参数可使网络在

信号统计中适应各种缓慢变化, 而且后面的主分量的附加计算在每一迭代中仅需要线性量级 $O(n)$ 的乘法运算, 在实验中, 取5维的 $M=20$ 数据点集用到此算法。从协方差矩阵中获

取的主分量与用此算法通过 194 次迭代算出的主分量之间的均方距离可达到 0.34×10^{-3} 。

Chen 和 Lin [14] 将反馈连接的概念推广到从训练数据中同时提取 M 主分量, 是 APEX 算法的级数计算的逆过程。网络的前向计算与 (22) 式相同。此外, 正交权 C 的训练法则与 (24) 式相同。主分量向量 $\{w_1, \dots, w_M\}$ 的学习规则修正成为:

$$\Delta w_i = \alpha \{B_i [y_m x - y_m^2 w] - A_i w_i\} \quad (26)$$

$$\text{这里 } A_i = \begin{cases} 0 & i = 1 \\ \sum_{j=1}^{i-1} w_j w_j^T & i = 2, 3, \dots, N \end{cases} \quad (27)$$

$$B_i = I_i - A_i \quad (28)$$

矩阵 A_i 和 B_i 在训练过程中是正交的, Chen 和 Lin 证明: 当“anti-Hebbian”权 B_i 收敛于零向量时, 权向量 $\{w_1, \dots, w_m\}$ 收敛于 M 主分量。

5) 主分量稳定性估计 Xu 和 Yuille [15] 已经发表了主分量估计中的稳定性问题。为了考虑训练集以外的数据, 他们首先引入了包括分布内的数据和排诉训集以外的数据的二元域。因为此函数是不可微的, 所以他们提出了基于 Gibb 分布的权函数, 用来表征数据样本分布的偏离程度。通过建立极小化的能量函数 $J(x, w)$, 它的梯度下降学习法则成为:

$$w = w - \alpha D_{\beta} \eta(x, w) \Delta J(x, w) \quad (29)$$

这里 $D_{\beta} \eta(x, w)$ 是权函数, 定义如下:

$$D_{\beta} \eta(x, w) = (1 + \exp[\beta(J(x, w) - \eta)])^{-1} \quad (30)$$

这有效地减小了分布以外点的影响, 分布以外数据点是指有大的能量函数值 $J(x, w)$ 的点。参数 β 是给定退化参数, 开始取值为小值然后到无穷大, η 作为确定外部分布区域参数。就能量函数 $J(x, w)$ 的选择而言, 许多 Hebb 学习法则可用某些能量函数的梯度法表示。

6) PCA 算法讨论 这些学习法则在从数据集中计算 M 主分量方面较传统特征分解技术有许多优点。如果 $M \ll N$, 这些迭代技术在计算上更有效; 其次, 由于它们的迭代特点, 这些算法将适应输入流的各种缓慢变化; 第三个优点是不需要再分配比特数存储数据或它的高阶统计; 最后, 如果需要再分配, 那么使用迭代学习法则将使其计算更有效。

应用神经网络的这些 PCA 算法可分为二类: 一类为再估计(重估计)算法, 它仅使用前反馈连接; 另一类为去相关算法, 此算法既有前反馈连接也有反馈连接 [16]。GHA 即是前一类例子。学习法则 (17) 式也可用下式表示

$$w_j(t+1) = w_j(t) + \alpha(t) * w_j(t) * [w(t) - \hat{x}(t)] \quad (31)$$

这里 $\hat{x}(t)$ 是重估因子, 定义为:

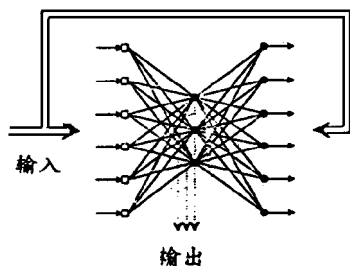
$$\hat{x}(t) = \sum_{k=0}^L w_k(t) * y_k(t) \quad (32)$$

在学习过程处理数据之前, 这种网络的连续输出必须学习从输入减去最初分量的估计值所得的主分量差。相反, APEX 是去相关算法。“anti-Hebbian”连接去除连续输出的相关性, 从而计算主分量差。

最近, 学者们有意于将以上方法扩展到一个新的领域, 即非线性 PCA 网络中, 其中神经元的模型中加入 S 形“活性”函数。用此模型, 可从数据中获取高阶统计; 然而, 基向量的结果失去了相互之间的正交性。虽然非线性 PCA 已经成功地应用到正弦信号的分离上, 但由于失去了正交性, 所以它们在图象压缩中的实用性受到限制。

7) 自联相变换编码 Cottell 和 Munro [17] 和 Cottell, Munro 和 Zipser [18] 已经在图

象压缩中应用了具体的编程方法。他们认为图象压缩问题是编码问题,其中网络必须通过“窄”的信道执行恒等映射。例如,当隐层用16节点时,输入和输出层由 8×8 节点组成。在这种方案中,每层均充分连接。网络的训练集由大量随机选择的子图象组成。每幅子图象同时出现在输入和输出层上,层之间的权按照反向传播的算法修正。图11为此网络结构。



在实时图象实验中,在可接受的主观失真水平下,图11 自相关反向传播网络 Cottrell 和Mumro 得出的压缩比为1:8。同时线性网络产生的结果于非线性网络的相近。在另一个实验中,每一个隐层节点仅连接顶层和底层的 4×4 区域。这样,减少到原来连接的11% 仅使均方差增加30%。

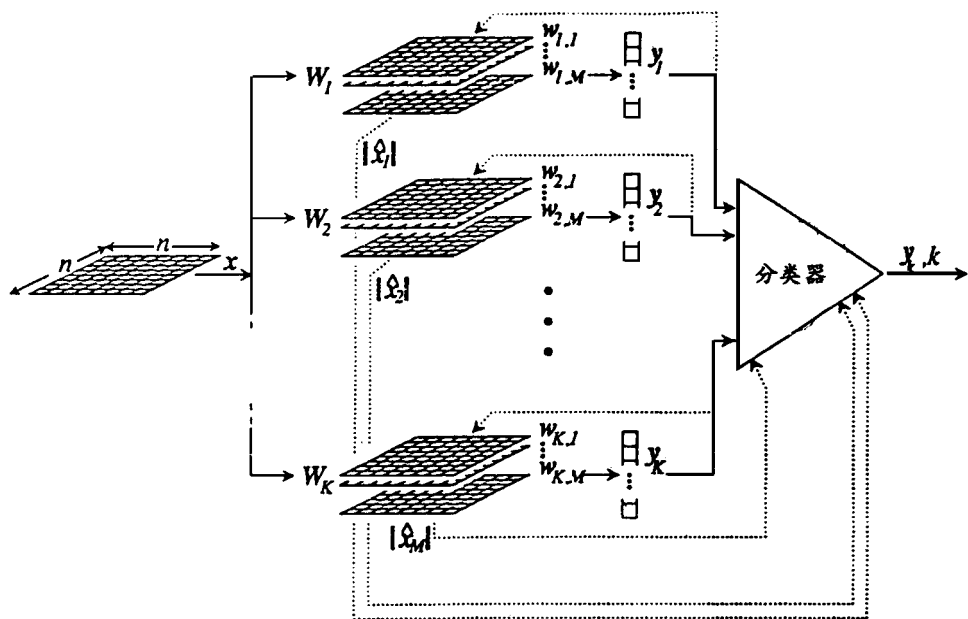


图12 OAL 结构框图,输入是 $n \times n$ 象素方块, K -变换 W_i 由大小为 $n \times n$ 块的 M 个基图象组成,且输出 M -维向量 y_i ,传送的系数是基于投影向量模 \hat{x}_i 最大的子空间分类器选择的系数。

8) 自适应变换编码 已经提出的自适应网络,它将Hebb学习和竞争学习在拓扑序图中结合起来。此网络适应于许多以自组织方式分布的混合数据[19, 20]。图12示此系统编码阶段的模结构。此系统由大量独立的模式组成,它的输出被子空间分类器选择。每模式由 $n \times n$ 维的 M 图象基组成,它定义了可逆的线性变换。每个图象基与输入图象块作内积,导出每一模式的 M 系数,作为 M 维向量 y_i 。每一模式对应着一种输入数据。类型的选择和因此被传送的系数向量以及它的类索引一起由子空间分类器决定。这种选择是建立在投影向量模 \hat{x} 是最大的类别上。投影向量 \hat{x} 采用系数向量的逆变换来计算。图象的解码应用相同的变换集。类索引常用于选择逆变换类和计算结果重构图象块 \hat{x} 。

4 应用神经网络的向量量化

1) 向量量化 量化处理是将信号 $X(n)$ 映射成一系列 K 离散信息,对于第 k 个信息,存

在一对门限值 t_k, t_{k+1} , 输出值 q_k , 使得 $t_k < q_k \leq t_{k+1}$ 。对于给定的一组量化值, 理想的门限值与这些值等距。量化数据的概念可以由标量或一维数据推广到任意维的数据向量, 在非输出水平, 向量量化(VQ)使用表征向量集(一维情况)或矩阵(二维情况)。集称为“码本”, 每一元素称作“码字”。由距离矩阵定义的决定面代替门限值, 特别是应用码字的欧化距离。向量量化较标量量化的优点是可以得到相邻象素之间的较高相关性。即使对于无记忆系统, 向量编码较标量在理论上易于改进性能。

在编码期间, 图象被分隔成大小为 $n \times n$ 像素块。对于每一块可以得到在距离矩阵下最近的码本的码字及其序号。重建时, 应用同一码本进行简单的查阅操作就可得到重建的图象。计算码本的标准方法是应用LBG的算法[21]。开始时, 在输入空间根据每一块最近的编码字将其分类, 然后每一编码字由它的最终分类的平均值所替换, 迭代继续下去直至得到最小的可接受误差。这种算法在训练过程中将均方差减到最低, 当LBG算法收敛到局部最小时, 并不保证达到全局最小。另外, 此算法对初始码本非常敏感, 尤其是此算法收敛缓慢, 因为在每一迭代中它都需要对整个码本进行彻底的搜寻。

2) 自组织特征映射算法 Kohonen 的自组织特征映射(SOFM)[22]已经成为在向量量化中应用网络模型来解决码本设计问题的大量研究的基础。Kohonen 引入拓扑图特征类序概念。在许多算法中如 K -均值, 每个向量 x 被分类, 只有获胜选定类在迭代过程中参加修正[23], 在 SOFM 算法中向量 x 不仅获胜选定的类要学习, 而且它的相邻类也要按如下规则学习。

在训练集中对每一个向量 x :

(1) 根据下式分类 x

$$\text{若 } x - w_i = \min_j x - w_j \quad \text{则 } x \in c_i \quad (33)$$

(2) 根据下式修正 w_j

$$\text{若 } w_j(t+1) = \begin{cases} w_j(t) + \alpha(t)[x - w_j(t)], & w_j \in N(w_i, t) \\ w_j(t), & w_j \notin N(c_i, t) \end{cases} \quad (34)$$

这里 w 是特征向量, α 是取值为 $0 < \alpha < 1$ 的学习参数, $N(c_i, t)$ 是在 t 时刻获胜类 c_i 的相邻类组成的集合, 类的特征向量 w_i 收敛于类均值, 相邻类是以对类的拓扑序的某种距离测量来确定的。例如, 如果在二维坐标格上对类进行排序, 某类的相邻类是指与此类的欧氏距离小于某特定的门限值类的集合。起初, 在训练中, 相邻类可以相当大, 即为类数的一半或更多, 随着训练进行, 相邻类的大小不断缩小直到最后仅包括一类为止。在训练中, 参数 α 也缩小到小值(即 0.01)以达到算法的较好收敛状态。

3) SOFM 算法的特征 SOFM 算法有许多重要的特征, 这些特征使它更适合用作向量量化的码本发生器[5]。

(1) 特征向量集是原输入空间的优化逼近。

(2) 特征向量在特征图上是按拓扑排序的, 因此向量之间的相关性随他们之间距离减小而增加。

(3) 特征映射的密度与输入分布密度相对应, 因而带有高概率密度的区域比带有低密度区域求解更容易。

4) SOFM 与 LBG 算法的比较 用于码本设计的 LBG 算法与 SOFM 算法很相近[24], 事实上, 对于同一邻域范围 $N(c_i) = \{c_i\}$ 来说, LBG 算法与 SOFM 算法是等价的。通过将向

量 w_i 定为类的均值, LBG 算法使均方差(MSE)的失真在类内减低到最小,

$$D_{MSE}(x, w_i) = E(x - w_i)^2, x \in c_i \quad (35)$$

均值是在 n 个训练样本之后以一批众数来测算的。为了在每训练样本一次之后学习 w_i 向量, 可采用基于减少方差失真(35)式的梯度下降方法。学习规则为

$$w_i(t+1) = w_i(t) - \frac{1}{2} \alpha \nabla D_{MSE}(x, w_i(t)) = w_i(t) + \alpha(t) [x - w_i(t)] \quad (36)$$

在 $N(c_i) = \{c_i\}$ 时, 它等价于(35)式。因此二者都计算类均值, 从而有最小的MSE码本。

许多学者[25]~[28]已经成功地用SOFM算法来产生VQ码本, 并论证了许多较经典LBG算法的优越之处。其优越之处包括对码本初始化的低敏性, 较低失真率和收敛的快速性。已经证明了码字或权收敛于类的均值[29, 30]。就平均失真被降低而言, 结果码本是理想的。

5) 地址预测向量量化 在码本设计中应用SOFM算法的另一个优点在VQ的变化中显示出来, 称为地址预测向量量化(APVQ)。此方法应用有序的码本, 码本中相邻码字在一定意义上是相关的, 图像中相邻码字间的相关性和邻块间的相关性可用来建立DPCM编码器, 其中输入信号是码字地址。这种技术可提高编码效率, 并允许地址码有(失真)误差。与标准VQ相比当SOFM算法协同APVQ用来设计码本时, 编码图象仅需小于37%的比特数。

6) 有限态向量量化(FSVQ) SOFM也已成功应用到有限态向量量化(FSVQ)方案中[31], 在FSVQ方案中, 从现有状态 S_n 的状态码本中挑选码字索引 i_n 。现有状态是前状态 S_{n-1} 和前索引 i_{n-1} 的函数。如果状态过渡函数是下一个输入较好的预测器, 那么每一状态码本就要比无记忆量化器所需的码本少得多。Liu和Yun[31]应用单个超码本的子集作为状态码本来代替每一状态的分离码本。应用SOFM算法计算超码本而且在拓扑映射之内状态码本是相邻的。在图象中, 现态码本仅是码字周围到与现输入块相邻的前输入块的相邻块。他们发现对于给定的比特率, 同时应用SOFM和FSVQ可使SNR较无记忆VQ增加4.2dB; 另一方面对于同样的失真, 比特率减少一半以上。

7) 学习向量量化 SOFM算法计算一组向量 $\{w_1, w_2, \dots, w_{1c}\}$, 这组向量用作向量量化的码字。学习向量量化是监督学习算法, 如果有一组标记的训练数据, 它可用来修正码本[22]。对一个输入向量 x , 设最近的码字索引为 i , 且设 j 为输入向量的类脚标。码字 w_i 可按如下修正。

$$(1) \text{ 若 } i = j \text{ 则 } w_i(t+1) = w_i(t) + \alpha(t) [x - w_i(t)] \quad (37)$$

$$(2) \text{ 若 } i \neq j \text{ 则 } w_i(t+1) = w_i(t) - \alpha(t) [x - w_i(t)] \quad (38)$$

这里 α 是取值为 $0 < \alpha < 1$ 的学习率参数。如果说学习率参数以0.1开始, 然后随着每次迭代单调下降, 在一定次数的迭代后, 码本很典型地收敛训练终止。

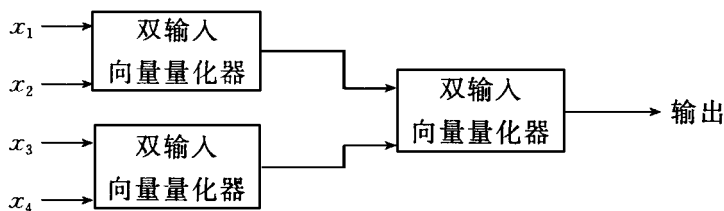


图13 基于双输入向量量化器的双层等级向量量化器

8) 等级向量量化 传统的VQ 编码的一个缺点是在编码阶段所须的运算负荷, 因为对每一输入向量都需要它对整个码本进行彻底搜寻。一个可供选择的方法是以等级方式串联许多编码器, 此方法对编码速度及精度进行权衡[30, 32]。此方法如图13所示。在每样本为2比特或每个输入向量为8比特时, 对于整个搜索空间为32码字编码器, 每一阶段仅需要 $2^4 = 16$ 个码字, 而有相同比特率的传统编码器则需 $2^8 = 256$ 个码字。另外, 当假定平移不变时, 每阶段仅需计算一个码本, 在一维信号计算机实验中, Luttrell[32]发现: 与类似单一阶段编码器相比, 这种等级编码器在失真方面仅增加0.05 dB。

另一降低搜寻时间的方法是在拓扑结构的有序码本中, 开发码字组织。在SOFM 算法的初始训练阶段, 由于初始大邻域函数, 每一输入向量修正许多码字。随着训练的进行, 邻域数目减少。一些学者已经提出了一种训练算法, 这种方法可使邻域保持不变, 但网络的大小却增加了[33, 34]。应用这种方法在多次迭代之后, 网络大小以量级方式成倍增长。新的码本插入到现存的码本之间并初始化它的相邻均值, 这在训练过程中使计算大幅度减少, 而仍能产生拓扑结构的有序码本。Luttrell[33]已经成功地使用此技术来产生SAR 图象的码本。

结果码本可视为网络的等级[35, 34] 它的每一级是前一级的一半大小, 见图14; 图中底层包含网络的所有码字, 同时每个下一层包含前一层的码字的子集。对一个输入向量, 搜索从高层次开始并发现获胜的码字。然后搜索获胜码字在下一层的相邻码字, 不断重复这一过程直到发现底层的获胜码字, 即整个码本搜索结束。搜索是复杂性的对数级, 因而此算法比全搜索更有效。在图象实验中, Truong[34]发现当这一技术相对于全局搜索奇迹般地降低训练和编码时间时, 失真的增加是很小的, 最多减少SNR 0.5 dB。

Lee 和 Peterson[29]已经开发了一种方法, 这种方法可以很明显地改变权之间的结构和邻域关系, 通过允许产生新权, 合并和删除现有权, 并在拓扑映射中合并轴线。

9) 频率敏感竞争学习 LBG 算法的一个疑难问题就是在码本中码字的使用频率可以极不规则, 一些码字未能使用。在频率敏感竞争学习(FSCL) 中[36~ 38], 通过训练期间增加码字的获胜数函数来权化训练中的码字距离。在这种方式下, 获胜多次的码字, 即过度使用的码字, 将在训练过程中不作为修正的选择。就欧氏距离度量而言, 修正的距离测算可以下式表示。

$$D_{FSCL}(X, w_i) = F(u_i) \|X - w_i\|^2 \quad (39)$$

这里 u_i 是码字 i 获胜的次数, $F(*)$ 是获胜数的增加函数称为“公证函数”。

Ahalt 等和 Krishnamurthy 等[36~ 38]已应用 FSCL 算法设计图象的VQ 码本。两种“公证函数”被采用, 一个等于获胜数目:

$$F(u) = u \quad (40)$$

另一函数随着训练进行减少到1。

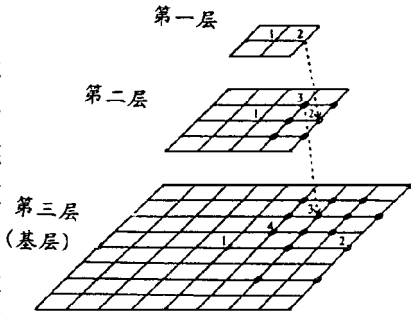


图14 节点1是根节点, 节点2是根节点类中与输入向量距离最小的节点, 节点3是第二层节点2相邻的获胜节点, 节点4是第三层节点3相邻的获胜节点, 基层是整体码本 and 所有节点(码字)所处位置。

$$F(u) = u^{e^{-t/T}} \quad (41)$$

这里 t 是训练迭代变量, T 是训练进行一半时 $F(u) = u^{e^{-1}}$ 时所选择的值。他们发现“公证函数”对结果码本有显著的影响。应用 (40) 式较 LBG 码本可使 MSE 增加, 而使用 (41) 式其 MSE 与 LBG 码本的几乎相同。

5 总结讨论

应用神经网络解决图象压缩问题的深入研究已经产生了一些很好的结果, 由神经网络的自身特征决定了它很适合图象数据处理, 人工神经网络的特征与人类自己的许多视觉系统特征相类似, 这些特征包括大量的并行结构、高度的相关性、贮存经验知识的倾向和自我组织能力。相比之下, 图象数据处理的传统算法一直是基于信息处理的序列模型, 它更适合于连续信息如语言的处理。已经证明神经网络方法在图象压缩应用中较传统方法要好。由于神经网络的非线性特性, 可用它来设计预测编码的非线性预测器, 已经证明经反向传播算法训练的多层感知预测器相对于线性预测器能提高预测效果。基于高阶统计模型和回归模型的其它网络方法也同样证明较线性预测器有所改进。

Hebbian 学习已经构成了许多迭代方法的基础, 这些迭代方法用来取得图象数据的主分量, 以使用作块变换编码的图象基。GHA 算法与 APE 算法已经证明可收敛到 M 主分量, 这些算法以及它们的变化形式较样本自协方差矩阵的标准特征分解有许多优点。仅需要开始几个主分量, 就可显著地节省计算。此外, 当不需要贮存所有数据或它的高阶统计时, 存储需求也减少了。通过用一个非线性变换, 自相关网络也已经成功地使用到压缩图象数据方面。

SOFM 算法已经以许多方式应用到 VQ 领域, 它能以自组织方式形成有序的拓扑特征映射, 这种能力使得它较标准的 LBG 算法在 VQ 码本的形成中有许多优点。已经发现 SOFM 算法有如下的优点: 对初始条件的低敏感性、收敛速度快以及产生较低平均失真码本的能力。同时, 等级 VQ 和预测 VQ 方法已经利用码本的序特征来降低查寻的复杂性和减小失真。

随着各种各样应用神经网络方法解决图象压缩问题的出现, 很迫切地需要对新的方法进行综合评价, 这种评价将需要共同的训练集及训练集以外的共同测试集。神经网络算法的性能要与类似的“经典”方法比较, 这种比较应该建立在失真度量, 包括 MSE (均方差) 和感知失真度, 比特率和编码、解码及训练复杂性基础上。

尽管仍有一些问题还需进一步研究, 但许多图象压缩的神经网络方法显示了很好的前景。然而, 只有这些方法真正能实现其平行结构, 这些方法的全部潜力才得以实现。在以上研究的大部分实现方法是基于序列模型模拟的。随着许多神经网络结构补充的 VLSI 的发展, 训练和编码的速度就均会大大地提高。

参考文献

- 1 Rumelhart D. E., McClelland J. L., Eds. Parallel Distributed Processing. MIT Press: Cambridge, MA. 1986
- 2 IIPMANN R. P.. IEEE ASSP Magazine, Apr 1987, Vol 4: 4~22, Apr 1987
- 3 Kohonen T., Self-Organization and Associative Memory. Berlin: Springer Verlag, 3rd ed, 1988

- 4 Hush D. R. , Horne B. G. . IEEE SPM magazine, Jan 1993 Vol 10: 8~ 39
- 5 Haykin S. , Neural Networks: A Comprehensive Foundation New York: Macmillan, 1994
- 6 Dianat S. A. , Nasrabadi N. M. , Venkataraman S. . In Proc IEEE Int Conf Acoust, SSP, Toronto, Canada, May 1991, 2793~ 2796
- 7 Li J. , Manikopoulos C. N. . Nonlinear predictor in image coding with DPCM. Electron Lett, Aug 1990, Vol 26: 1357~ 1359
- 8 Manikopoulos C. N. . IEEE Proc Oct 1992, Vol 139: 501~ 507
- 9 Pao Y. H. , Adaptive Pattern Recognition and Neural Networks Reading MA: Addison-Wesley, 1989
- 10 Haykin S. , Li J. . To be published in IEEE Trans, Signal Processing, 1995
- 11 Haykin S. . Adaptive Filter Theory. Englewood Cliffs, NJ: Prentice Hall, 2nd ed , 1991
- 12 Hebb D. O. . The Organization of Behavior 1949, New York: Wiley
- 13 Sanger T. D. . Neural Networks, 1989, Vol 2: 459~ 473
- 14 Chen H. , Liu R. . In Proc IEEE Int Conf ASSP 92', San Francisco CA, Mar 1992, : II 283~ 296
- 15 Xu L. , Yuille A. . Tech Rep: 92~ 93, Feb 1992, Harvard Robotics Lab
- 16 Oja E. . Subspace Methods of Pattern Recognition Letchworth UK: Research Studies Press 1983
- 17 Cottrell G. W. , Munro P. . In SPIE Vol 1001 Visual Commun, and Image Process, 88': 1070~ 1077
- 18 Cottrell G. W. , Munro P. , Zipser Z. . In 9th Annu Conf of the Cognitive Soc , July 1987, : 462~ 473
- 19 Dony R. D. , Haykin S. . In Proc IEEE Int Conf ASSP Apr 1993, : I 609~ 612, Minneapolis, MN.
- 20 To be published in IEEE Trans Image Processing, 1995
- 21 Linde Y. , Buzo A. , Grey R. M. . IEEE Trans commun, Jan 1980, Vol C- 28: 84~ 95
- 22 Kohonen T. . Proc IEEE Sept 1990, Vol 78: 1464~ 1480
- 23 Duda R. O. , Hart P. E. . Pattern Classification and Scene Analysis New York: Wiley, 1973
- 24 Luttrell S. P. . In IEEE Conf On Neural Networks 1989, : 495~ 498, Washington DC.
- 25 Nasrabadi N. M. , Feng Y. . In SPIE Vol 1001 Visual Commun, and Image Process, 88' 1988, : 207~ 213
- 26 In Proc IEEE Int Conf on Neural Networks, July 1988, I: 101~ 105, San Diego, CA.
- 27 McAuliffe J. D. , Atlas L. E. , Rivera C. . In Proc IEEE Int Conf ASSP 90', Apr 1990, II: 2293 ~ 2296, Albuquerque, NM.
- 28 Manohar M. , Tilton J. C. . In Neural Networks for Perception, Vol I Human and Machine Perception, San Diego CA: Academic, 1992, : 345~ 367
- 29 Lee T. C. , Peterson A. M. . IEEE J. On Selected Areas in Commun , Oct 1990, Vol 8: 1458~ 1471
- 30 Li J. , Manikopoulos C. N. . In SPIE Vol 1199, Visual Commun, and Image Process, IV 1989, : 1046 ~ 1055
- 31 Liu H. , Yun D. Y. . In Proc Of the Int Workshop on Applications of Neural Networks to Telecommunications Hillsdale, NJ: Lawrence Erlbaum Assoc 1993: 176~ 182
- 32 Luttrell S. P. . IEEE Proc (London), Vol 136(part I), 1989, : 405~ 413
- 33 In Proc IGARSS Sept 1988, : 1231~ 1238, Edinburgh, Scotland
- 34 Truong K. K. . In Proc IEEE Int Conf ASSP May 1991, : 2789~ 2792 Toronto, Canada
- 35 Truong K. K. , Mersereau R. M. . In Proc IEEE Int Conf ASSP Albuquerque, NM. Apr 1990, : 2289~ 2292

- 36 A halt S C , Chen P , Krishnamurthy A. K . In Proc IEEE Int Joint Conf on Neural Networks 1989, 1: 169~ 175
- 37 A halt S C , Krishnamurthy A. K , Chen K , Melton D. E . Competitive learning algorithms for vector quantization Neural Networks, 1990, Vol 3: 277~ 290
- 38 IEEE J. on Selected Areas in Commun Vol 8: 1449~ 14

Neural Network Apply to Image Compression

Liu Chunyang Liang Dequn Song Huansheng Wu Gengshi
(Xi'an Jiaotong University, Xi'an 710049)

Abstract

This paper presents methods of neural network as signal processing tools for image compression. They are well suited to the problem of image compression due to their massively parallel and distributed architecture. Their characteristics are analogous to some of the features of our own visual system, which allow us to process visual information with much ease. For example, multilayer perceptions can be used as nonlinear predictors in differential pulse-code modulation (DPCM). Such predictors have been shown to increase the predictive gain relative to a linear predictor. Another active area of research is in the application of Hebbian learning to the extraction of principal components, which are the basis vectors for the optimal linear Karhunen-Loeve transform (KLT). These learning algorithms are iterative, have some computational advantages over standard eigendecomposition techniques, and can be made to adapt to changes in the input signal. Yet another model the self-organizing feature map (SOFM), have been used with a great deal of success in the design of codebooks for vector quantization (VQ). The resulting codebooks are less sensitive to initial conditions than standard LBG algorithm, and the topological ordering of the entries can be exploited to further increase efficiency and reduce computational complexity.