

南京航空航天大学

研究生硕士学位论文

姓 名： 章 雅 珍

专业名称： 通 信 与 电 子 系 统

研究方向： 图 象 分 析 与 计 算 机 视 觉

指导教师： 成 瑜 教 授
王 金 香 副 教 授

一九九九年三月

论文题目

自适应 BP 网络压缩编码

研究生姓名：章雅珍

专业：通信与电子系统

研究方向：图象分析与计算机视觉

指导教师：成渝 教授

授予学位：工学硕士

答辩时间：1999 年 3 月

入学时间：1996 年 9 月

毕业时间：1999 年 3 月

一九九九年三月

A Dissertation for the application of Master's Degree

An adaptive compression coding with BP Neural Networks

Prepared by

Zhang Yazhen

Under the Supervision of

Prof. Cheng Yu

**DEPARTMENT OF ELECTRICAL ENGINEERING
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS**

Mar.1999

摘要

图像压缩编码是图像传输, 存储中迫切需要解决而又未能完全解决的问题。本文提出了一种利用自适应分类人工神经网络来进行图像压缩的编码方法。

神经网络具有大规模并行处理及分布式信息存储的优点, 有良好的自适应性、自组织性和容错性, 有很强的学习功能、联想记忆功能, 引入图像数据压缩中的 BP 网络有很强的数据压缩能力。首先, 将图像分成独立的 (8×8) 图像子块来减少运算量。由于当 BP 网络直接提供数据压缩时, 其网络通用性差, 所以采用分类的办法来分别进行压缩。同时, 采用人的视觉特性曲线来进行分类以充分利用人的视觉特性, 提高重建图像的主观质量。在分类过程中, 还利用了图像的局部统计特性和图像子块的活动特性, 这样, 进一步提高重建图像的主客观质量。为了消除图像分类中的错误, 进行了分类提升。由于图像常有大面积相似区域, 对相邻图像子块中的相似子块进行合并, 不但可以提高压缩比, 而且可以减少压缩时间。

通过这些方法, 用分类训练后的神经网络来进行编码, 通用性能较单个神经网络进行压缩的方法有了较大改善, 而重建图像的主客观质量也有了很大提高。

关键词: 神经网络, 图像压缩, 自适应, 分类编码, 反向传播 BP 算法, 压缩率, 峰值信噪比

ABSTRACT

Picture compression coding is a problem that is urgently need but not overcome completely in the field of picture transmission and picture memory. In this thesis, an adaptive classified and artificial neural network coding method for still picture is proposed, where human visual properties and local picture statistics block activities are utilized.

Neural network models are introduced in the compression of picture data, with intrinsic parallelism, distributive memory, error-tolerance property, and strong study ability, BP neural networks have the ability of direct data compression. In order to reduce the computational effort, a image is divided into (8×8) pixels image blocks. But because of lack of the generalization property, the directly BP neural network data compressing method must be combined with block classified coding method. The human visual properties are introduced into the work of classifying the blocks to improve the subjective quality. During the classifying, the picture local statistic properties and block active properties are also utilized to get advanced quality. When the classifying appears error, the grade of classifying is rise to eliminate the error. Finally, similar block convergence will fully utilize the large similar areas that are often appeared in pictures.

After all the above-mentioned means, the trained BP neural networks, when are used in picture coding, have a better generalization property and the reconstructed picture have a significant improved performance of subjective and objective quality.

Key words: neural network (NN), picture compression, adaptive, classified coding, back-propagation (BP) algorithm, compression ratio, peak signal to noise ratio(PSNR)

目 录

第一章 绪论-----	1
§ 第一节 图像压缩编码的发展概述-----	1
§ 第二节 本文设计方案的特点-----	6
第二章 数据压缩理论及压缩性能评价-----	9
§ 第一节 数据压缩的理论极限-----	9
§ 第二节 数据压缩的基本途径-----	15
§ 第三节 图像压缩系统的性能评价-----	19
第三章 图像块的自适应分类-----	22
§ 第一节 自适应分类的基本原理-----	22
§ 第二节 自适应分类算法-----	23
§ 第三节 分类的实验结果-----	25
§ 第四节 分类的改进-----	28
第四章 BP 神经网络编码器-----	31
§ 第一节 BP 网络压缩编解码器的工作原理---	31
§ 第二节 BP 网络的改进措施-----	36
§ 第三节 自适应分类的 BP 网络编码器-----	9

附录：一类 BP 网络编码器的连接权值	43
第五章 实验结果和分析	46
§ 第一节 BP 网络压缩编码器的压缩结果	46
§ 第二节 实验结果分析	49
结束语	50
致谢	53
参考文献	54

第一章 绪 言

人类社会已进入信息时代,而信息的本质,则要求交流和传播。采用数字技术(或系统)具有许多优越性,但也使数据量大增。信息时代带来了“信息爆炸”。数据压缩的作用及其社会效益、经济效益将越来越明显。如果不进行数据压缩,则无论传输或存储都很难实用化。而数据压缩的好处就在于^[1]:

1. 较快地传输各种信源(降低信道占有费用)——时间域的压缩;
2. 在现有通信干线上开通更多的并行业务(如电视、传真、电话、电报、可视图文等)——频率域的压缩;
3. 降低发射机功率——能量域的压缩;
4. 紧缩数据存储容量(降低存储费用)——空间域的压缩。

图像作为一种高数据量的信息,如果不进行数据压缩,无论是存储,传输都很难实用化,但图像具有直观性、生动性等特点,因此很有必要对其进行数据压缩,使其实用化。

第一节 图像压缩编码的发展概述^[2]

电视信号数字化(线性 PCM)在 1948 年提出以来,迄今已有将近五十年的历史,当时即提出图像编码(频带压缩)的研究方向。在五十年代和六十年代,限于当时的计算机和电路实施的客观条件,反对预测法和亚取样/内插复原法进行过一些基础性的研究;对视觉特性也做了一些极为有限而宝贵的工作。

1969 年在美国举行首届“图像编码会议”(Picture Coding Symposium,简称 PCS),表明图像编码以独立的学科跻身于学术界。

时至今日,图像编码技术已获得众多令人瞩目的成就,在经典的编码技术方面,生命力强的编码技术在获得广泛应用的同时,仍在不断地发展。例如,研究最早的预测法,仍在显示其“青春活力”而广泛应用。其比较注目的研究进展有:

- 1) 我国学者利用“大预测误差”的频度较小为参考，以图像主观质量为准则，研究了电视信号的场内预测编码^[3]；
- 2) 与自适应预测编码相对应，提出了图像的组合信源模型；
- 3) 在 CCITT H.261 建议中采用最简单的帧间预测^[4]；
- 4) 信息保持型（Lossless）压缩编码在 JPEG（Joint Photographic Expert Group）建议中^[5]仍采用预测法；
- 5) 预测法基本思想（图 1-1）也同样适用于 JPEG 建议^[5]的扩展系统以及模型法图像编码。

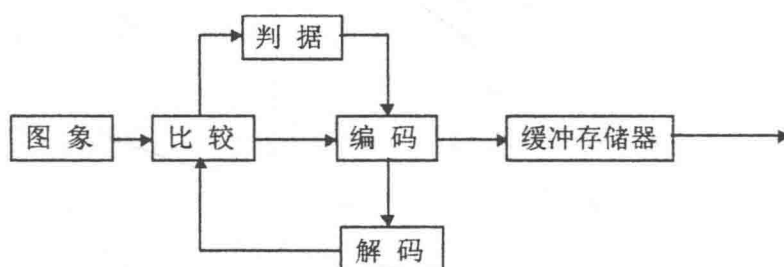


图 1-1 预测法的基本思想

又如七十年代初开始迅速发展并成为研究热点的变换法编码方法，尤其是变换法编码中的 DCT(Discrete Cosine Transform)编码^[6]，由于其性能非常接近 KLT 而得到最广泛的应用。近几年的主要成果有：

- 1) 确定了 DCT 像素块的最佳尺寸为 (8×8) 像素^[5, 7]；
- 2) 制定了对 DCT 逆变换（IDCT）算法精度的要求^[8]；
- 3) 提出了对电视电话/会议电视和常规电视的 DCT 变换域“视觉域值矩阵”之参考数据^[5, 9]；
- 4) 为了消除“块效应”（block_effect）而采用简单的数字滤波器^[7]。

更深入的研究有：

- 1) 将像素块的形状由矩形推广到任意形状^[10]；
- 2) 开始研究能够反映视觉特性的非正交的 Gabor 变换^[11]。

其他如亚取样/内插复原法，运动估值^[12]作为运动补偿的技术，方块截尾编码^[13]，矢量量化^[14]等方法都有相当的发展。

1988 年在图像编码的发展史上是极为重要的一年。四十年研究成果的集中表现在：电视电话/会议电视的 CCITT H.261 建议^[7]和静止图像压缩编码的 JPEG 建议^[5]的原理框图已经基本确立。

为了解决“多分辨率”(multiresolution)图像的压缩编码,近年来一些学者致力于几种新编码方法(子带(sub-band)编码、塔形(pyramid)编码,以及小波(wavelet)编码)的研究。但从原理上说,它们仍属于线性处理,因而仍可归入经典图像编码方法。

1988 年以来有较多发展的各种现代图像编码方法的显著特点是:它们突破了经典图像编码方法所依据的信源编码理论的框架,在同样压缩比条件下,现代方法比起经典方法处理后的重建图像在主观质量方面有显著的改进。其主要有下列三大类图像编码方法:

一. 三维线框(wireframe)模型法(简称模型法,包括二维线框模型),这是把计算机视觉和计算机图形学中的方法应用到图像编码,并取得显著的研究进展。这类方法进一步划分为两小类:有先验(a priori)知识的和无先验知识的。有先验知识的大多应用于电视电话等领域,景物中的主要对象(object)是人物的头肩部分(head-and-shoulder),可预先在发送端和接收端事先约好分别设置两个相同的三维线框模型;无先验知识的如我国的卢朝阳等提出的基于 Delaunay 三角形化(triangulation)的“表面描述”(surface description)之图像编码方法^[15]。在这类编码方法中,怎样描述编码误差是一个未能很好解决的问题。

二. 分形(fractal)用于图像编码。分形是十多年前 Mandelbrot 提出的几何学新概念。自然界中的大多数景物比起欧几里德几何图形要复杂得多。如树木在分形理论中由于采用新的几何定义,其维数为 1.5。分形最显著的特点是自相似性(self-similarity),即景物中的任何一小部分的形状都与较大部分的形状极其相似。分形用于图像编码比较有效的方法是 1984 年 Barnsley 提出的迭代函数系统(Iterated Function System,简称 IFS)^[16],这种方法可获得很高的压缩比,但它的主要困难在于如何找到恰当的码。

三. 人工神经网络(Artificial Neural Networks,简称 ANN)用于

图像编码^[17]。目前将神经网络理论和技术引入图像编码领域的研究发展相当快。人脑在接受视觉传感器传来的大量图像信息后,能够迅速作出反应,并在脑海里重现这些信息,这不仅与人脑的巨大信息存储能力有关,还说明人脑具有较强的特征提取功能,也就是数据压缩能力。人工神经网络虽然只是人脑的简单模拟与抽象,但具有许多与人脑相似的信息处理能力,自然具有较强的数据压缩能力。目前用于图像编码的神经网络模型有近十种。

在对图像所包含的信息来分析,它既有随机性,也有规律性,期望用一种图像模型来描述自然界千奇百怪的景物图像是不现实的。由于图像内容变化的随机性,即使用几种模型组合也难以对图像有效地分割和数学描述。鉴于神经网络在解决黑箱问题上的特别有效性,可以用神经网络学习图像中规律性的东西,将人们对图像的了解,即先验知识,巧妙地结合到系统的构造中,对图像中随机性的东西,通过神经网络自适应机制,如结构自适应、学习率参数的变化和连接强度的变化等进行调整,可以利用神经网络的特点对图像信息进行有效的分解、表征和编码,从而取得传统方法无法比拟的结果。

从理论上讲,编码问题可以归纳为映射和优化问题。映射可以是线性的,也可以是非线性的。不少神经网络完成的恰好是一种由输入空间到输出空间的数学映射,另外,优化类的神经网络在组合优化问题中已充分展示了优越性。

如前所述可知,对于随机性很强的图像信号来说,很难找到一种与不同统计特性的图像信号完全相同的变换(而如果所选的基矢量与图像信号的协方差矩阵的特征矢量只是接近,则变化域内的图像信号相关性会明显下降),因此变换矩阵中基矢量的选择不是固定的,这就给最佳变换编码的实现带来困难。神经网络的强映射能力和非线性特性,使它可以学习具有接近输入信号特征空间基的能力,因此用来解决最佳变换的实现是有效的。

由于神经网络模型大多具有相当强的模式识别与模式分类能力,这也给图像编码方案中模式分类问题的解决提供了一个强有力的工具,如图像的矢量量化编码就是一个典型的模式分类问题,神经网络

算法可使之得到有效的解决。

同时,神经网络还有较强的容错性和联想记忆功能,任何局部的损坏不会影响整体结果,这一特性有助于对有噪图像的数据压缩及对压缩后信息不全图像的恢复。另外,神经网络的大规模并行处理能力,为神经网络图像压缩编码的实时实现创造了条件。

常用于图像压缩编码的神经网络主要有 Hopfield 网络、BP 网络、Kohonen 自组织神经网络、模拟退火 (SA), CPN 网络等。其中有些神经网络模型能提供直接的数据压缩,而有些是间接参与数据压缩。

Hopfield 神经网络是一种优化类的神经网络,用于图像编码领域有:实现了变换编码中的最优化问题并用硬件实现;实现了 BTC;实现了矢量量化编码;实现了求非完全 DCT 逆矩阵等。但由于其异步计算,计算复杂度高,运算结果与初始值的选取和动态演化过程密切相关,其实际应用还有困难。

Kohonen 自组织神经网络是一种自组织类神经网络,具有较好的拓扑聚类特性,用于图像压缩编码主要是矢量量化器的码书设计。

SA 学习算法是一种新的优化算法,将全局搜索和局部搜索融为一体。主要用于矢量量化技术中,但运算时间较长。

BP 网络是目前最为常用的一种神经网络模型,可直接提供数据压缩。在文献[18]中已经证明,三层和三层以上的 BP 神经网络能以任意精度逼近任何线性和非线性映射,图像编码与 BP 神经网络在数学本质上的这种意义上的同构是 BP 神经网络能直接进行压缩的原因。

N.Sonebara 等人于 1989 年探讨了三层前馈神经网络的通用性与训练图像数目及迭代次数的关系,还探讨了隐节点输出值量化和初始权值的选择对重建图像质量的影响,为了控制网络规模,将图像分成 8×8 的子块,每一子块分别送到相应的子网络进行并行处理,因此,学习速度快。G.W.Cottrell 与 N.Sonehara 等人提出的三层神经网络虽然可以直接提供数据压缩,但通用性较差,对非训练图像,编码质量显著下降,边缘失真大,训练图像的编码质量也不尽如人意,为此,提出了许多改进方案^[19, 20, 21, 22, 23]。

Z.He 与 H.Li, S.A.Dianat, 张伟等人将多层前馈神经网络用于图

像非线性预测编码, 其性能优于线性预测器^[24]。M.Arozullah 等人提出了一种对称结构的五层神经网络模型, 采用扩展的 BP 嵌套训练算法, 压缩比可达 64: 1, 国内的张秀明等人也对五层网络模型的图像压缩进行了有意义的探讨^[25], 王卫等人也提出了一种对称塔形神经网络 (SPNN), 都取得了很好的压缩效果。

BP 网络除了具有强影射能力外, 还具有较强的模式分类能力。F.Mekuria 等人利用感知机模型设计了一种快速码书搜索算法, M.N.Pour 等人将多层前馈神经网络用于矢量量化编码, 取得了相当好的性能, 但网络隐节点数随矢量维数与码书的增大呈指数律增长。

1991 年 S.Marsi 等人^[20]提出的一种改进结构的图像压缩方法, 根据图象块的活性参数, 将其分成六类, 每一类送到相应的 BP 网络进行训练, 不同的 BP 网络根据图象块的活动特性取不同的隐节点数。对训练集外的 Lenna 获得峰值信噪比 (PSNR) 为 25.8dB, 比特率为 0.38bpp 的压缩效果。

第二节 本文设计方案的特点

本文是在 S.Marsi 和 R.Kohno 等人提出的方法^[20, 19]的基础上提出的一种改进方法。我们所提出的自适应分类神经网络编码方法, 是根据人眼的对比灵敏度特性曲线和图像块的活性参数作为图像子块的分类的标准的, 每一类送到相对应的三层 BP (Back-propagation) 网络进行训练, 不同的 BP 网络, 根据分类类别的不同取不同的隐节点数。自适应分类神经网络编码方法具有以下几个特点:

一: 本方案采用的是 BP 网络压缩编码器。由于人脑具有较强的特征提取功能, 也就是数据压缩能力, 人工神经网络虽然只是人脑的简单模拟与抽象, 但具有许多与人脑相似的信息处理能力, 具有大规模并行处理及分布式信息存储的优势, 有良好的自适应性、自组织性和容错性, 有很强的学习功能、联想记忆功能, 具有较强的数据压缩功能。BP 网络由于其结构简单、直观, 具有很强的映射能力, 可直

接提供数据压缩功能。

用于数据压缩三层 BP 神经网络包括输入层、隐含层和输出层，隐含层的节点数要少于输入节点数，输入节点数与输出节点数相同，整个网络结构完全是对称的。学习时，图像数据既送到输入层作为输入，又送到输出层作为教师信号，所使用的学习算法是 BP 算法。网络训练好后，输入层到隐含层为网络的编码器，从隐含层到输出层为网络的译码器，它是一种结构压缩法。其隐含层输出为压缩后的数据，输出层的输出为解码复原的数据。

BP 网络虽然训练时间较长，但学习后在线运算量却很小，并且具有并行计算结构。

二：本方案利用了人眼的对比灵敏度特性和图像子块的活动性来对图像子块进行分类。这是将人眼的对比灵敏度特性引入神经网络压缩编码的一种新的尝试。将人眼的对比灵敏度特性结合到图像子块的分类中，就较好地利用了人眼的对比灵敏度特性。如在平坦区，由于图像子块的灰度值变化范围不大，肉眼对如此小的灰度变化感觉上无多少变化，就可用图像子块的平均灰度值来代替整个图像块中的灰度值。同时，根据图像子块的活动参数，还对灰度变化较大的边沿区再细分成四类（详见第三章）。

这样，既提高了图像压缩比，又提高了图像的主客观质量。

三：在图像自适应分块中，还采用了分类提升的办法来减少其分类时的错误，即当重建图像子块和原始图像子块间的误差超过一定的误差门限时，将图像子块分类时的类别号提升一个等级（详见第三章）。这样可提高图像的压缩质量。由于图像中经常会出现大面积的相似区域，在编码过程中，还对相邻的相似子块进行合并，这样，就可以大大提高其压缩比。

四：在 BP 神经网络设计过程中，针对图像压缩编码时输入、输出节点数较一般 BP 神经网络分类器的数目大，输出层输出值的大小的精度要求比分类器时的输出要小的多，因此在调试过程中，采用了多种改进措施（详见第三章）来提高 BP 神经网络的收敛速度和收敛精度。

整个方案的流程图如图 1-2 所示。

本文第一章介绍了压缩编码的基本原理和基本途径，以及重建图像质量的评价方法；第二章介绍了根据人眼视觉特性和图像子块活动参数特性进行分类编码及其结果；第三章介绍了 BP 神经网络编码器的工作原理，调试特点和实验结果；第四章主要是应用本方案对一些图像的压缩结果及一些改进意见。整个方案采用 Borland C++ 语言来仿真实现。

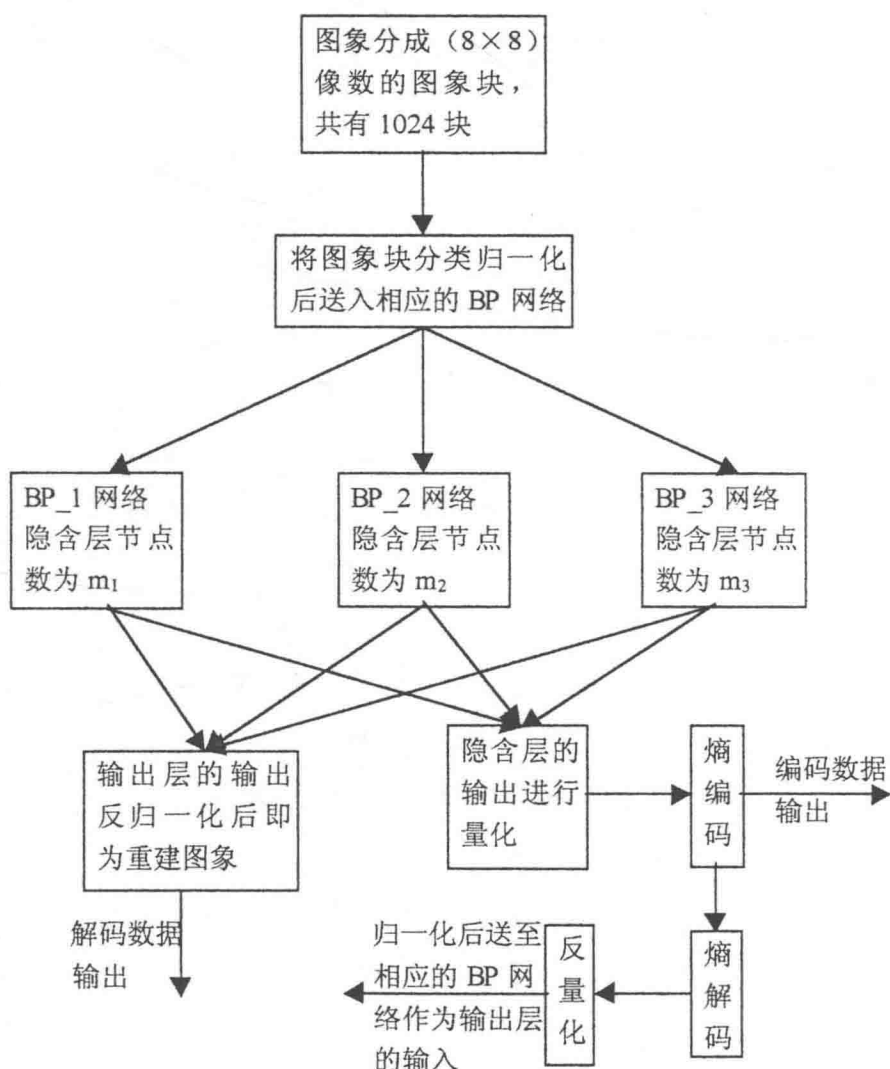


图 1-2 编解码的流程框图

第二章 数据压缩理论及压缩性能评价^[1, 6, 26]

经典的数据压缩理论是建立在信息论基础上的, 1948年由香侬所创立的信息论, 对于数据压缩这门学科有着极其重要的指导意义: 它一方面给出了数据压缩的理论极限, 同时又指明了数据压缩的技术途径。

第一节 数据压缩的理论极限

作为香侬信息论研究的对象——信息, 被假设为由一系列的随机变量所代表它们往往用随机出现的符号所表示, 我们称输出这些符号集的源为“信源”。研究中, 不考虑这个源的内部结构及发生符号的机理, 只研究这些符号集的属性。那么, 由信号源输出的随机符号, 如果其取值于某一连续区间, 就叫做连续信源; 如果其取值于某一离散集合, 就叫做离散信源; 如果随机符号的一部分取值于连续区间, 另一部分取值于离散集合, 则称之为混合信源。一般地说, 信源发出的消息是一个随机过程, 它是时间与空间的函数。

我们用大写字母(如 X)表示随机变量, 小写字母(如 t)表示随机变量的一个实现。于是一个离散信源的输出可用序列集合

$$\{X_t; t=0, \pm 1, \pm 2, \dots\}$$

来表示, 集合中每个元素取自字母表(有限符号集合)

$$A_m = \{a_1, a_2, \dots, a_m\} \quad (2.1-1)$$

中的一个, 而字母表中的元素叫作字母或字符。若一张表中含有 m 个不同的字母, 就说该表的大小为 m 。

若取 t 为有限数 n , 则信源又可用 n 维随机矢量来表示, 即

$$X = (X_1, X_2, \dots, X_n), \quad X \in A_m^n \quad (2.1-2)$$

这里的 A_m^n 是 A_m 中各元素的 n 重笛卡儿乘积(积集), 总共有 m^n 种可能组合。其中每一个都叫作长为 n 的源字。

用 $P_n(X)$ 表示 n 维随机矢量

$$X_t = (X_{t1}, X_{t2}, \dots, X_{tn}), \quad X_t \in A_m^n$$

的概率，并记

$$t+k \sim (t_1+k, t_2+k, \dots, t_n+k)$$

若对任意的整数 k 与 n ，所有的 $X \in A_m^n$ 都满足

$$P_{t+k}(X) = P_t(X) \quad (2.1-3)$$

则称此信源为平稳信源，此时式(2.1-3)的下标 t 可省去。若对任意 $X \in A_m^n$ ，又有关系式

$$P(X) = \prod_{t=1}^m P(X_t) \quad X_t \in A_m \quad (2.1-4)$$

成立，则此平稳信源又称为离散无记忆信源（常记为 d.m.s.）。

我们记字符 a_j 出现的概率为

$$P(a_j) = p_j, \quad j=1, 2, \dots, m$$

那么（按概率的公理化定义）必须有

$$0 \leq p_j \leq 1, \quad \sum_{j=1}^m p_j = 1 \quad (2.1-5)$$

则字符 a_j 出现的自信息量定义为

$$I(a_j) = -\log p_j \quad (2.1-6)$$

$I(a_j)$ 的含义实际是：随机变量 X 取值为 a_j 时所携带信息的度量。

在人们日常生活当中，如果收到一封电报，接到一个电话，或得到一组数据，从信息论的观点看，只能称之为“消息”（Message），而不能说收到“信息”（Information）。不过，信息寓于消息之中，而且同样一个消息会带有不相等的信息量。这是因为每个消息送到用户那里，对用户所提供的收益和价值，将因事而异。比方同样收到一封仪器实验成功的电报，如果原先已经预计到成功的可能性是 99%，那么它所带来的信息量是微小的，好象有“不出所料”之感。但是如果预计成功的可能性不大，比如只有 1% 的可能性，而这时收到成功的电报，就会给人们提供极大的信息量，似乎有“十分意外”的味道。此时对人们心理上的鼓励作用，在技术上敢于肯定的积极态度，都表明该消息所含有的信息量是很多的。

显然，用户在收到消息之前，是处于某种不确定状态之中；而