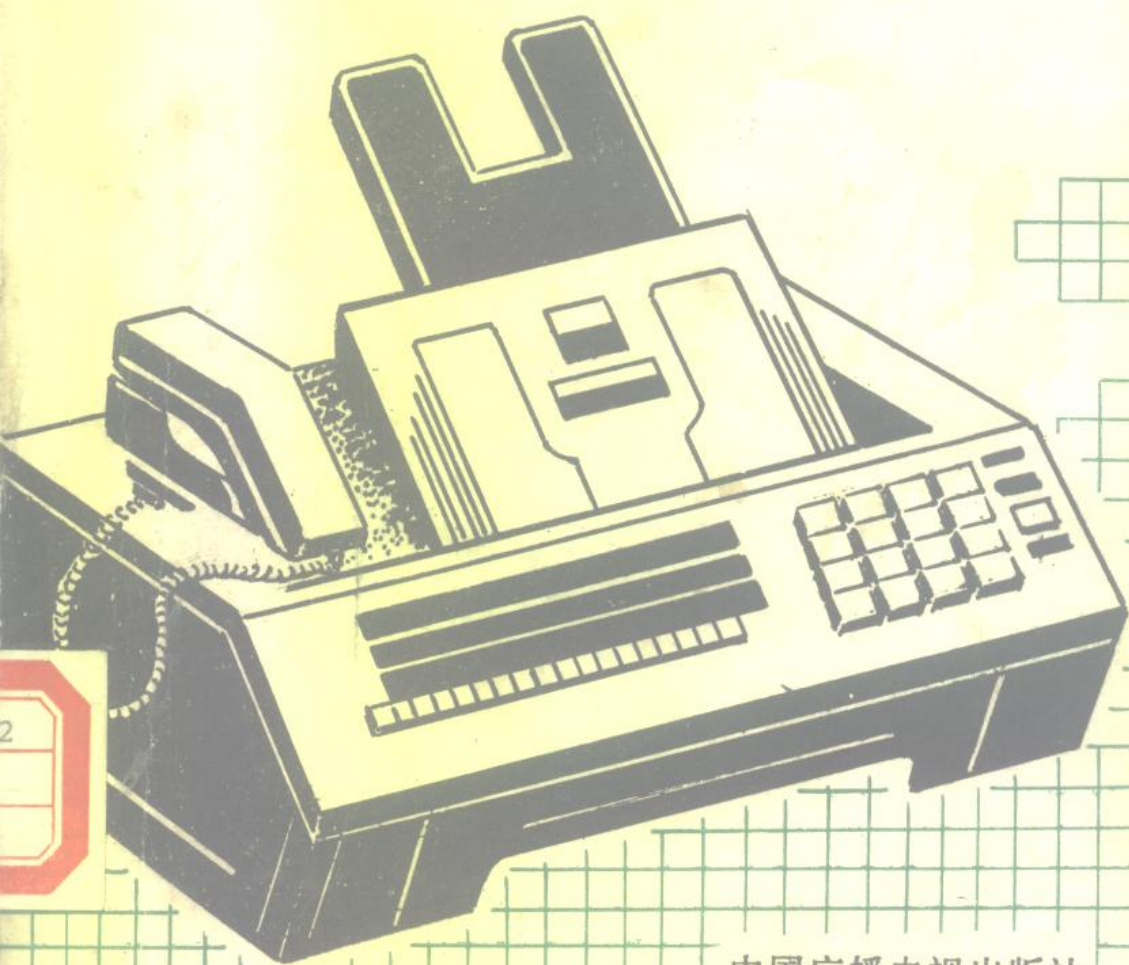
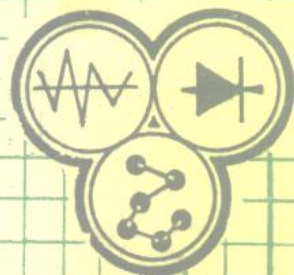


传真三类机电路分析 及维护调整

周子耀 编著



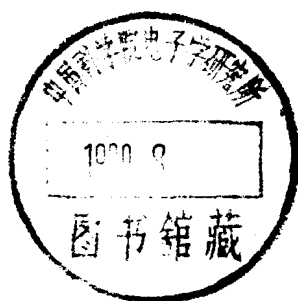
中国广播电视出版社



73.6622
352

传真三类机电路分析及维护调整

周子耀 编



中国广播电视出版社

9010218

传真三类机电路分析及维护调整

周子耀 编

中国广播电视出版社出版
(北京复外广播电影电视部灰楼 邮政编码100866)

河北望都新风印刷厂印刷
新华书店总店北京发行所经销

787×1092毫米 16开 14印张 280(千)字

1990年5月第1版 1990年5月第1次印刷

印数: 1—8000册 定价: 7.20元

ISBN 7—5043—0405—0/TN·39

前 言

文件(黑白)传真机是一种现代化通信设备。我国自60年代广泛使用传真机以来,传真通信在公用电话网上传送了大量文字和图像。进入80年代,特别是最近几年,我国的传真通信得到了迅速发展,用传真三类机代替了传真一、二类机。目前,传真三类机在电信、党政机关、国防、交通、气象、海洋、新闻、金融、工业及商业等部门获得广泛使用,有的部门还租用了电信网,使电话、传真及数据处理一体化。这样,传真机就成为办公室自动化必不可少的终端设备。

在我国,传真三类机受到越来越多的用户欢迎,其主要原因是:传输率高、接收质量好、管理功能强、通信费用低、操作简单及易于与计算机互联等优点。我国1980年只进口15台传真三类机,而到1980年进口两万多台,预计今后进口更多。日本是生产传真机的大国,约占全球总数的99%以上,型号也越来越多,但所生产的传真三类机在CCITT建议的基础上对功能有所扩展,在设计中注意到了进入ISDN的一些规则。

我国引入的三类机,大部分为日本冲电气及松下公司的不同型号的产品,特别是OKI FAX7700E及UF2EXC最多。因为用户对传真三类机的原理了解不够,而对软件设置知道的更少,在使用中不尽一致,影响了使用效果,当传真机出现故障时更是束手无策。为此,有必要编写一本资料讲述传真机基本原理、通信规程并结合目前使用最多的7700E详细分析硬件工作原理,为培养维修人员提供参考依据。尽管传真三类机型号很多,但硬件结构差异不大。通过对7700E机硬件的分析,也会对其他型号传真机电路原理的理解有极大帮助。

根据以上情况,本书第一章、第二章简要讲解了传真三类机方面的一般基础知识;第三章、第四章和第五章详细地分析了CCITT传真三类机建议T.4和T.30;第六章是本书的重点,用大量的篇幅详细解说了各单元电路工作原理及7700E机的使用、维修。对各电路测试点列出了参考波形图,为维护人员提供了方便。

本书的读者对象主要是大专院校通信专业师生、机务维修和使用人员,通信和办公室自动化方面的技术人员。

本书初稿已在1985年完成,因当时手头资料少,现在出版进行大的改动已很困难,请读者阅读时参考有关资料。

本书在编写过程中得到中办一局领导及研究所的技术人员、北京电子专科学校有关教师的大力帮助,并提出了宝贵的意见和建议。本书出版得到了中共安徽省委办公厅一局刘良寿及北京电子专科学校李兴民、修秀荣、崔健等老师的支持,在此一并表示衷心感谢。

由于编者水平有限,书中难免出现错误,敬请广大读者批评指出。

编 者

1989年10月于北京

目 录

第一章 传真三类机概貌

- 第一节 概述..... (1)
- 第二节 传真三类机的传输系统..... (2)

第二章 数字化频带压缩基础

- 第一节 传真信号的数字化..... (5)
- 第二节 传真信号的统计特性..... (6)
- 第三节 信源编码的基本原理..... (9)

第三章 传真信号编码

- 第一节 一维改进的霍夫曼码 (MHC) (18)
- 第二节 二维编码..... (22)
- 第三节 编码的扩充不压缩模..... (29)

第四章 调制解调器 (MODEM)

- 第一节 V.27ter调制解调器..... (30)
- 第二节 V.29调制解调器..... (33)
- 第三节 V.21调制解调器..... (38)

附录..... (39)

第五章 传输控制规程

- 第一节 操作方法的分类..... (41)
- 第二节 传真呼叫的时间顺序及传真呼叫描述..... (41)
- 第三节 传真过程用的单音信号方式..... (48)
- 第四节 传真过程用的二进制信号方式..... (53)

第六章 OKIFAX7700收发两用机

- 第一节 概述..... (69)
- 第二节 技术参数及主要性能..... (74)
- 第三节 发送与接收系统..... (78)
- 第四节 主要印刷电路板..... (94)
- 第五节 通信管理机能..... (187)
- 第六节 传真机的联接与设置..... (197)
- 第七节 检查调整及测试..... (200)
- 第八节 使用维护及维修..... (206)

后记..... (207)

附录一 信号流程总图..... (210)

附录二 7700传真机使用的集成电路型号表..... (211)

附录三 缩写表..... (214)

第一章 传真三类机概貌

第一节 概 述

传真三类机是以数字信号进行处理的通信设备,所以又称为数字传真机。它以图像的统计特性为理论基础,对传真信号进行数字化编码,以消减传真信号的多余度,从而提高了传输效率。具体的编码方案和码字的组成形式很多,但原则上均按概率分布分配长短不同的码字。例如:游程长度编码(Run—Length Coding简称RL编码),给概率大(出现频次多)的游程分配较短的码字,而给概率小(出现的频次少)的游程分配较长的码字。

数字化是传真的发展方向,这是因为数字传真机的压缩比高,抗干扰能力强,而且可以采用分组交换技术,便于加密。另外,数据通信是今后社会信息传输的方向,特别是信息的存储、处理、交换和加密等均需要信息数字化,传真机的发展必须考虑这一前景。为此,1980年5月,国际电报电话咨询委员会(简称CCITT)完成了文件传真三类机的国际标准化,为三类机的推广使用铺平了道路。

直到目前,三类机的性能不断完善,如1989年日本冲电气推出的OF-17传真三类机、带有保密接口的UF-200传真机以及现在问世的多快度、彩色、普通记录纸传真机等,都是最近的新产品。由于传真三类机功能强、用途广,所以它除了用于通信外,还用作办公室自动化中的终端设备,估计目前世界上至少有5000万台传真机在使用中,已成为世界范围内的普及型通信设备。例如,目前日本使用的2200万台传真机中,三类机约占90%。

我国国家标准局于1982年12月发布了传真三类机的国家标准,即GB3382—82。该标准是我国传真三类机在公用电话交换网和租用电路的互通技术标准,也是传真三类机设计和制反订维护标准的依据。

利用电子计算机对传真图像信息进行处理越来越广泛,但必须使其传真信号数字化。过来,传真信号数字化,又使传真机成了电子计算机的外部设备,用以输入图像和记录图像,从而使传真技术跨入了更加广泛的领域。

近几年来,三类机使用了微处理器,使其性能更加完善。微处理器在三类机中主要应用以下部分:

总机控制部分;

数据压缩及扩张部分(即编码及译码部分);

调制解调部分;

误码纠错重发部分;

线密度自动调整及信息存储部分;

步进电机的驱动部分等。

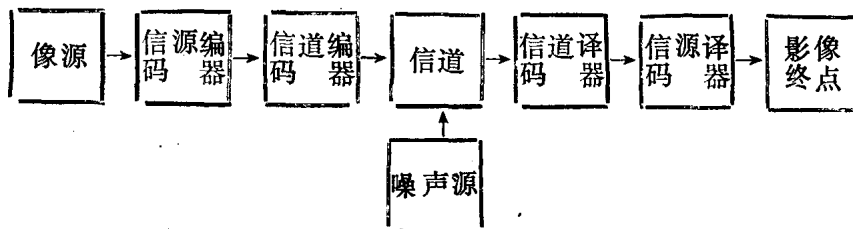
随着大规模集成电路和数据通信的发展,1985年CCITT对文件传真四类机已经作出规

定。规定它是一种减少文件信号多余度的，并优先考虑在公用数据网上使用的设备。它采用适于公用数据网的规程，并保证无误码接收文件。当采用适当调制解调方式时，也可在公用电话网上使用。

第二节 传真三类机传输系统

一、影像传输系统模型

传真是影像传输系统中的一部分，其方框图见〔图1.2.1〕



〔图1.2.1〕影像传输系统

由图可看出，影像的传输系统由八部分组成。

影像源——是由所要传输的图像经光电变换后的电信号信息源，

信源编码器——它是把信息源的数据经过适当的编码(如三类机采用的是MH或MR编码方式)，以达到消除多余度的目的，

信道编码器——它是把信源编码器送来的信息，变换成适于在线路上可靠传输的部件，

信道——指传输线路，可分为模拟信道和数字信道两大类，

噪声源——影像信号在信道传输系统中所受到的随机干扰，

信道译码器——对信道编码器的反演，

信源译码器——对信源编码的反演，

影像终点——影像源的重现。

二、传真三类机基本标准

根据我国GB3382——82国家标准等效采用CCITT的传真三类机建议。其主要内容是：

1、扫描轨迹

传真三类机是采用平面扫描方式，但它的扫描定义，仍按照一类传真机扫描轨迹，对其各像素顺序作规定，即发送机与接收机都以相同方向对报文进行扫描。我们从垂直平面上来看报文区，其扫描方向是从左到右，且随后的扫描线紧接在前一扫扫线之下，再从左到右……。

2、扫描线长

按规定传真三类机标准的扫描线长度为215mm。这是根据北美地区公文用纸规格(8 $\frac{1}{2}$ 英寸)制定的，比我国公文用纸GB148—59—A,规格(186mm×263mm)太。显然在扫描线长为215mm的传真机上传送宽为186mm的文件，将会有14%的幅面空白未有用来传送信息，这样对文件的传输会增加时间。经过对比，实际传输一页文件所增加的时间约7秒。

3、基本参数

基本参数是指在垂直方向扫描密度和水平方向扫描抽样密度。按规定标准有：

(一) 在垂直方向上, 标准扫描密度和选用的高扫描密度分别为3.85线/毫米 $\pm 1\%$ 及7.7线/毫米 $\pm 1\%$;

(二) 在215毫米 $\pm 1\%$ 长的标准扫描线上有1728个黑白像素, 还有共选用的255毫米 $\pm 1\%$ 长的扫描线有2048个黑白像素, 303毫米 $\pm 1\%$ 长的扫描线有2432个黑白像素。

并规定传真三类机可以接收最小尺寸为ISO A₄的输入文件(ISO—国际标准化组织, A₄幅纸尺寸为210mm $\times$ 297mm)。

一般在副扫描方向上的线密度采用3.85线/毫米, 若不能保证文件质量时, 可选用清晰度标准的7.7线/毫米, 由于清晰度提高使报文传送时间相应增加。

4、全编码扫描线的传输时间

全编码扫描线的确切定义, 要在了解编码方案后才能解释, 这里只定性说明其含意。

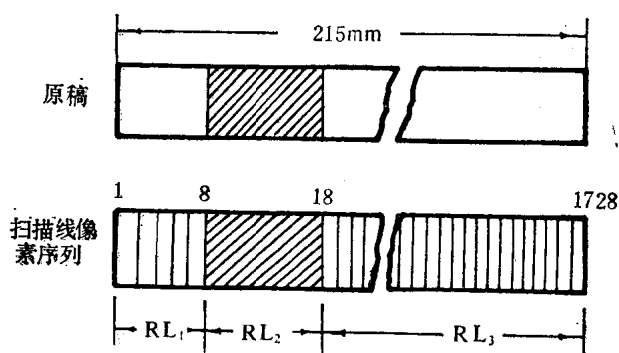
由传真扫描器拾取的一条扫描线的图像信号, 对应于原稿上被扫描的文字或空白图像的黑白像场。黑白像场对应的传真信号分别称为黑持续(或游程)长度和白持续长度, 这些交替出现的黑白持续长度就组成了一条扫描线上的传真信号。经抽样处理后, 这种持续长度以像素作为单位。这样一条扫描线上的传真信号就转换成一条像素序列, 如图1.2.2所示。

由图可看出, 原稿中被扫描的部分, 含有黑笔划及空白, 全扫描线上对应的有三个持续长度, 其中: RL_1 包括八个白像素, $RL_2 = 10$ 个黑像素, $RL_3 = 1710$ 个白像素。三类机的编码方案包含了含有黑白不同持续长度的编码, 即码字。这样一条扫描线的像素序列经编码之后, 用它所包含的黑白持续长度所对应的码字及必要的控制码字的总比特数来表示, 这就称为全编码扫描线。由此, 得出全编码扫描线的定义: 数据比特数加上所需的填充比特数以及线终码比特数的总和。

传送一全编码扫描线的编码数据所需要的时间, 称为全编码扫描线传输时间, 用字母T表示, 扫描线密度为7.7线/毫米时的T, 记作T_{7.7}, 扫描线为3.85线/毫米时的记作T_{3.85}。

由于三类传真机的编码方案是采用逐线编码方式。即处理编码数据时, 是以编码扫描线为单位的, 为保证三类机在某些部件不相同的情况下仍能互通, 比如扫描器、记录器的记录速度、编码器的编码效率等在有所差异的情况下, 由于信道均匀速率的要求, 必须对编码扫描线最小传输时间作出规定。三类机的标准扫描传输时间规定为20ms。此外, 还有选用的时间为0ms、5ms、10ms及40ms。

T为20ms时, 这反映了1980年制定国际标准时多数国家的传真机器件、部件的制造水平。T为40ms时, 只用于速度较低的传真机。若T小于20ms时, 并配有一定容量的缓冲存储器, 则文件的传输时间可减少。但当存储器容量的增加和传输时间T减少到某一程度



〔图1.2.2〕扫描线像素序列

时, 文件传输的时间减少就不够明显了。

为了适应三类机在技术上的实现和发展, 制造时可采用各种记录机构, 除规定 $T = 20\text{ms}$ 是标准的最小编码扫描线传输时间外, 还规定可选用的数值, 有 0ms 、 5ms 、 10ms 、 40ms 。这些不同的使用情况必须符合以下规定。

(一)、在扫描密度为 3.85线/毫米 (标准扫描密度)及 7.7线/毫米 (选用的高扫描密度)时, 编码扫描线的最小传输时间 T 应相同, 即 $T_{3.85} = T_{7.7}$ 。这时要求 T 为 10ms 的设备必须能降至 20ms , T 为 5ms 的设备必须能降至 10ms , 20ms , T 为 0ms 的设备必须能降至 5ms 、 10ms , 20ms 处理图像数据, 并能降至 40ms 。

(二)、在 $T_{7.7} = 1/2 T_{3.85}$ 时, 要求 $T = 10\text{ms}$ 的设备必须能强制退回到 20ms 来处理图像数据, 这是传真机的记录机构以连续记录两条相同的 7.7线/毫米 扫描密度的扫描线, 以获得扫描密度为 3.85线/毫米 的扫描线。

规定 $T_{3.85}$ 为 20ms 是标准值, 而 $T_{3.85}$ 为 10ms 、 40ms 是选用值。

除了对 T 值有一最小规定外, 还规定了 T 的最大值不超过 5秒 , 若在全编码扫描线超过 5秒 时, 收机必须拆线(不能接收)。这是因为, 即使发方以最低的数据信号速率 2400bit/s 进行传送, 用最大的编码扫描线时间时, 也不可能达到在 5秒 之内传输一条扫描线的编码, 因而把它作为一种非正常状态来处理。

一般来说, T 值小的三类机在某些情况下, 如文件的文字密度小, 传输电路的质量好, 它可用比较高的数据信号速率发送, 这样, 传输时间可以缩短。

5、编码方案

三类机采用的编码方案有两种: 一种为改进的霍夫曼(简称MHC)标准编码。三类机必须具备这种编码; 一种为可供选用的二维编码, 即改进的相对像素地址指定码(简称MREAD或MR)。有关编码的原理, 将在后面详述。

6、调制与解调

当传真三类机传送编码的传真信号时, 所用的调制解调器有两种: 一种是必备的标准调制解调器, 它是在公用电话交换网上使用的, 数据信号速率为 4800bit/s 、 2400bit/s , 要求这种调制解调器符合CCITT V.27ter建议的规定; 另一种是租用电路或高质量的交换电路时使用的调制解调器, 它符合CCITT V.29的规定, 传输速率为 $9600/7200\text{bit/s}$, 这是可供选择的方式, 目前的三类机都具备这种调制、解调方式。

7、发送机输出功率与接收机的输入功率

按标准规定, 发送机的发送功率电平应在 $-15\text{dB} \sim 0\text{dB}$ 内可调, 因为传真机用户安装地址不同, 因而各自的电平不同, 应按线路规定调到合适的输出电平。

按标准规定, 接收机的输入功率电平在 $0 \sim -43\text{dB}$ 内可调, 以保证接收副件的质量。

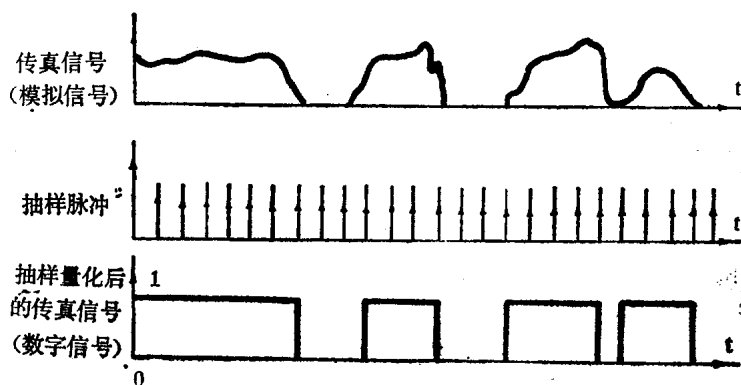
第二章 数字化频带压缩基础

黑白传真信号的多余度一般都相当大,通过数字化编码(也就是采用码化数据压缩方式)可以消减多余度,若只传输真正需要的有用信号,那么,所传输的数据量必然会大大减少。这样当频带一定时,消息传输时间可以减少;而当传输时间一定时,可以压缩所需的频带宽度。因为码化数据压缩方式实现频带压缩的效果比较明显。现在,码化数据压缩已是近年来在传真技术上的一个主要发展方向。

第一节 传真信号的数字化

经过扫描和光电变换后的传真信号,是一种连续模拟信号。在进行码化频带压缩之前,应将它转换成适于数字处理的数字信号。因而必须对这些模拟信号进行抽样和量化,使它在时间轴上和幅度轴上都能达到量化的程度。其作法是:将连续的模拟信号断续开来,变成一连串脉冲序列,离散地抽取信号的过程叫做抽样。抽出的各离散的信息单元,叫做样值。所谓样值,顾名思义是作为样值。而被抽出的各信息单元,必须能够代表被抽取信号的全体。对于连续的信号进行抽样的结果,若不能正确代表原信息时,可以把抽样间隔缩短(或者说提高抽样速率),以便抽出较多的样值。但是抽样时间间隔缩短到超过必要值以上时,在传输上仅仅增加带宽,既不必要又不经济。因而,有一个适合的抽样间隔是最合理的。经过证明找出了这个合适的抽样间隔,这就是抽样定理。这个定理的表述形式是:“来自信源的连续波形 $f(t)$,如果不包含 $f_{\max}$ 赫以上的频率成分,则该波形可由每隔 $1/2f_{\max}$ 秒的抽样值来唯一决定”。也可以表述成:“设信源包含的最高频率 $f_{\max}$ 赫,如果用至少为 $2f_{\max}$ 赫的速率抽出信息单元脉冲序列,则抽出的脉冲完全包含原来的全部信息”,

量化就是在表示某个量的过程中,按照一定的规律进行取舍,从而使该量表示成一定的标准量值。对于文件传真(黑白传真),所传送的原稿一段只有黑白突变的二值信号,但由于小孔畸变和黑白深浅本身的偏差,使经过光电变换后的电信号,对于这样的信号进行量化一般都取二值。〔图2.1.1〕是传真信号的抽样及量化情况。



〔图2.1.1〕 传真信号的抽样和量化

设该传真信号的最高频率(最高图像频率)为 $f_{\max}$ 赫。

抽样脉冲频率取 $2f_{\max}$ ，即抽样脉冲的间隔 $T = 1/2f_{\max}$ 秒。

这里必须说明，对于文件传真机，通常都对光电变换电路输出的传真信号先进行整形，这实际上起到了在幅度上进行了量化。所以，在实践中常先进行量化而后抽样。

第二节 传真信号的统计特性

采用码化数据压缩方式的数字传真信号，是利用传真原稿的统计特性来谋取多余度的削减。但从实用观点出发，实际利用的统计特性主要有：白像元与黑像元发生概率的差异，相邻元素之间的相关性以及统计量因图片不同而存在的差异等。

下面列出几种中文铅印文件的统计规律，如表2—1所示。

表2—1 各种概率

传真原稿 类别 模型		中 文 铅 印 文 件						
		NO1	NO2	NO3	NO4	NO5	NO6	TO7
维	P(B)	0.0695	0.0435	0.0615	0.0489	0.0556	0.0695	0.0449
	P(W)	0.9305	0.9565	0.9853	0.9511	0.9444	0.9305	0.9551
	P(B/B)	0.7114	0.5544	0.6253	0.6289	0.5428	0.5642	0.4812
	P(W/B)	0.2886	0.4456	0.3747	0.3711	0.4572	0.4358	0.5188
	P(W/W)	0.9784	0.9797	0.9754	0.9809	0.9731	0.9674	0.9756
	P(B/W)	0.0216	0.0203	0.0246	0.0191	0.0269	0.0326	0.0244

表中：P(B)——为黑像元出现的概率；

P(W)——为白像元出现的概率；

P(B/B)——为在同一扫描线上的某像元为黑时，其相邻像元也出现黑的条件概率，通常称之为状态迁移概率；同理，P(W/B)、P(W/W)、P(B/W)分别代表其余三种状态迁移概率。

NO1、NO2、NO3、NO4、NO5、NO6、NO7分别代表七种中文铅印文件样张，每种样张的文字内容相同，文件尺寸为 $194 \times 270\text{mm}$ ，文内无大标题，整页满满排印。

NO1：字样为三号字体，字高约5mm，行间间距约为4.5mm；

NO2：字样为三号仿宋体，字高约5mm，行间间距约为4.5mm；

NO3：字样为四号仿宋体，字高约4.5mm，行间间距约为4.5mm；

NO4：字样为楷体，字高约4mm，行间间距约为4.5mm；

NO5：字样为新四号字体，字高约4mm，行间间距4mm；

NO6：字样为新四号宋体，字高约4mm，行间间距约为2.5mm

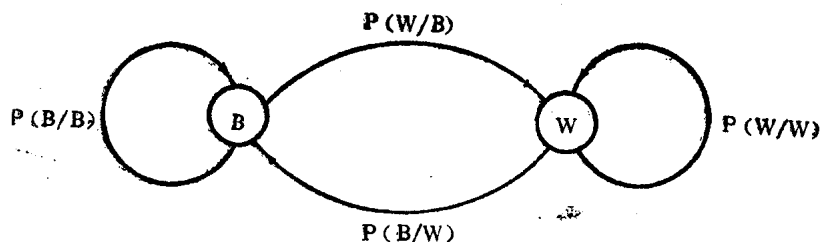
NO7：字样为小四号仿宋体，字高约4mm，行间间距约为4mm。

由表2—1可见，黑像元出现的概率P(B)小于10%，甚至不到5%，大部分为白像元；白像元出现的概率P(W)大于90%，一般文件传真都有这个性质。由白像元变为黑像元出现的概率P(B/W)，比起白像元单独出现的概率P(W)小得多，也比黑像元单独出现的概率P(B)小；由黑像元变为白像元出现的概率P(W/B)，比起白像元单独

出现的概率 $P(W)$ 也小得多,但比黑像元单独出现的概率 $P(B)$ 大。这说明相邻像元之间存在着很强的相关性,不仅存在主扫描方向,也存在于副扫描方向。

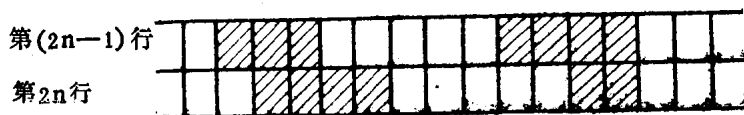
如表2—1所示,在同一条扫描线上的状态迁移概率有 $P(B/B)$ 、 $P(W/W)$ 、 $P(B/W)$ 四种,这种迁移可用一阶马尔可夫信源(一维状态)迁移图来表示,如〔图2.2.1〕所示。

考虑更多像元的相关性,如〔图2.2.2〕所示的情况,它是将两条扫描线综合在一起来考虑的,把相关的上下两个像元组成一个像元对,然后调查像元对之间的相关关系。



〔图2.2.1〕一阶马尔可夫信源(一维状态)迁移图

我们把二个像元分别取黑白两种状态,因此像元对有四种组合状态。于是在主扫描方向,相邻两像元对之间的状态迁移概率有16种,相邻两个像元对之间的状态迁移,如〔图2.2.3〕所示。



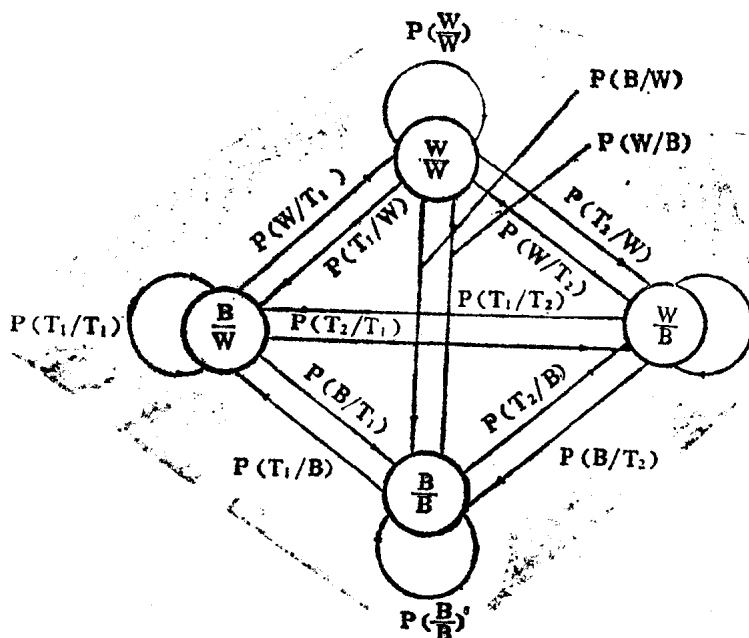
(a),像元对获得的方法

W	B	W	B
W	W	B	B

W T₁ T₂ B

(b) 像元对的状态

〔图2.2.2〕-像元对



〔图2.2.3〕二阶马尔可夫信源(二维)状态迁移图

我们利用上述的统计特性对黑白持续长度编码,可大量压缩数据量。如果对出现概率最大的持续长度用最短的码字表示,而对出现概率最小的持续长度用最长的码字表示,这样,出现的各种概率不同,分别分配不同的码字。这就是说,可以用最低限度的数据量来传送不同持续长度所包含的信息。我们把这种编码方法,叫作统计匹配编码,意思就是,码的长度是与概率分布相匹配的。这种编码方式,对数据的压缩率很高,因而称为最佳编码。

有关传真信号的一维游程编码和二维游程长度编码的数字模型,仍属于信息论的范畴,这里不再赘述。

第三节 信源编码基本原理

信源编码主要讨论如何将消息变成电信码,如何衡量电信码的好坏,什么样的电信码才算是最理想的编码等。

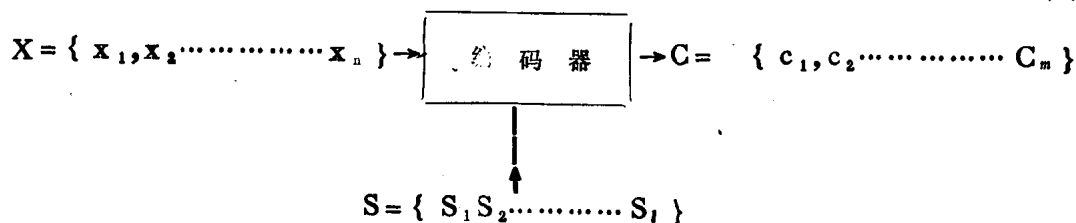
一、编码器的数学描述

设: $X = \{x_1, x_2, x_3, \dots, x_h\}$, $C = \{c_1, c_2, \dots, c_m\}$

$S = \{s_1, s_2, \dots, s_l\}$ 是三个有限集合。

X 代表信息集, x 代表信息单元; C 代表电码集, C 叫做码元(码字); S 代表符号集, 叫做构成 C 的符号。

编码器是这样一个部件: 如果在编码器的输入端送入一个 $x_i \in X$, 输出端将输出一个被指定与 x_i 对应的 $c_i \in C$, 而且所有的 C_i 均由 S 按规定的编码方法来构成。如图[2.3.1]所示。



[图2.3.1] 编码器

由图可以看出: 以符号 S (在我们以后讨论的编码器中, S 仅有 0、1 两种) 构成的电码, 建立 X 与 C 间的对应关系。

通过编码, 编码器输出一系列码字(电码) $c_i \in C$, 这一输出序列应是每一个码字都与一条原始消息相对应, 它们的对应顺序可规定为;

$C_i \longrightarrow x_i$

$C_i \longrightarrow x_i$

也就是说, 有一条消息就应有一个码字相对应, 而这个码字就代表这条消息, 不能代表其他消息。

如果 S 的所有符号均占用相同的时间, 则 S 称为同价符号空间。由同价符号构成的电码称为同价电码。反之, 如果 S 的符号长短不一, 则 S 称为非同价符号空间, 由长短不一的符号构成的电码, 叫做非同价电码。今后我们所讨论的范围属于同价电码。

二、编码要求

为了提高编码效率, 码字之间不应有间隔指示符号。为此, 编码器输出的电码应符

合以下要求:

1、任何一个电码都不是另外一个电码的延长——即电码为非延长电码,也就是说,我们不能在任何一个电码的后面添上一些符号作为另一个电码。

例如,由0,1两个符号构成的电码, $C = \{0, 10, 11, 100, 111\}$ 就是延长电码,因为100是10的延长,111是11的延长。而电码 $C = \{0, 10, 11\}$ 是非延长电码。

2、要求电码具有单义性。一般说来,单义性是我们对电码的不可缺少的要求,信源编码的目的,就是把消息变成码元,而且是平均长度为最短的单义电码。所谓单义电码,就是任意一个有限电码序列(有限长的一串C),它只能被唯一地分成为一个一个的码字。例如, $C = \{0, 10, 11\}$ 就是一种单义性电码,因为任意一串C,比如是100111000只能被分成10,0,11,10,0,0。而其他的分开方法都会有不是C的电码出现,如分成10,01,11,00,0时,其中01,00就不是电码集C中的电码。不是C中的电码就不能代表消息。所以,它只有唯一的划分方法,因而它是单义性电码,而且是符合前面提出的码元间不另外加间隔就能唯一分成一个一个码字。若 $C = \{0, 01, 10\}$ 不是单义性电码,因而如0010既可分成0,0,10;也可分成0,01,0。两种分法,各码都是电码集中的电码,都代表消息,但究竟哪种分法是正确消息还是不清楚的,这与我们对编码器的要求不相符合。

通过以上例子说明,非延长电码一定是单义性的。延长电码可能是单义性的,(例如 $C = \{0, 01\}$ 是延长电码,也是单义性电码),但延长电码又不一定都是单义性的。可见单义性电码是由非延长电码构成。

3、编码器输出的电码应是码字平均长度为最短的单义性电码。

我们把消息看成是一个概率空间X,其概率分布 $P(x)$,可写成如下对应关系:

$$\left\{ \begin{array}{l} x_1, x_2, \dots, x_n \\ P(x_1), P(x_2), \dots, P(x_n) \end{array} \right\} \quad \text{使码字 } C = \{c_1, c_2, \dots, c_n\}, \text{ 与 } X \text{ 产生对应关系。}$$

这样的编码称为样本编码。在数字传真机中应用的许多编码方案均属于样本编码。

样本编码的特点是 $P(C_i) = P(x_i)$ 。为简便起见,规定 c_i 对应 x_i ($i = 1, 2, \dots, n$)。则 $P(c_i) = P(x_i)$, $P(c_j) = P(x_j)$, 这就是说消息X的概率 $P(x_i)$ 也就是 c_i 码字的概率 $P(c_i)$, x_i 的概率 $P(x_i)$ 也是 c_i 的概率 $P(c_i)$ 。如果 $S = \{s_1, s_2, \dots, s_i\}$ 是一个同价符号空间,构成 c_i 码字的符号个数 n_i 可以代表 c_i 的长度。因此,同价电码C的码字平均长度为:

$$L = \sum_{i=1}^n P(c_i) \cdot n_i \dots \dots \dots (2-1)$$

现在的问题是如何构成一个单义性同价电码,在C与x顺序对应的情况下L最短。显然, $P(C_i)$ 大的码字,L应该小; $P(C_i)$ 小的码字,L可以大些。但是L最小能够小到什么程度呢?经理论证明,当下式

$$H(\quad) / \log_2 N \leq L < H(\quad) / \log_2 N + 1 \dots \dots (2-2)$$

成立时,L应该是同价单义电码的最短电码的平均长度。

式(2-2)称为样本编码定理。 $H(\quad)$ 为电码集的码字熵。 N 为符号集的元素个数, $H(\quad) / \log_2 N$ 称为下限,下限加1称为上限,这个对数以2为底是因为我们采用的信息单位是比特。

在我们所讨论的编码中, 符号集的元素只有0、1两个, 即 $N=2$, 所以 $\log_2 N = \log_2 2 = 1$ 。可见 $\log_2 N$ 的含义相当于每个符号的最大熵。

当 $L = H(\quad) / \log_2 N$ 时, 编码器的效率最高, 或者称为信源与编码相匹配。这时,
 $L = H(\quad)$ (因为 $\log_2 2 = 1$)

$$L = \sum_{i=1}^n P_i N_i, \text{ 式中的 } P_i \text{ 等于(2—1)式中的 } P(C_i)$$

$$H(\quad) = - \sum_{i=1}^n P_i \log_2 P_i. \text{ (概率空间的不肯定度, 或叫熵的公式)}$$

因为 $P_i N_i = -P_i \log_2 P_i$, 所以 $n_i = \log_2 P_i = 1 - \log_2 (1/P_i)$ (2—3)

上式为信源输出第 i 个消息的信息量, 由此可得出结论:

概率大的消息应编为短码; 概率小的消息, 应编为长码。

三、最佳编码方法

能够使得 L 为最短的编码方法, 叫最佳编码。用最佳编码方法编出的电码叫作最佳电码。下面以同价符号集 $S = \{0, 1\}$ 为例来说明最佳编码方法。

霍夫曼(Huffman)码

霍夫曼码称为理想最佳编码。这种码的特点是码字的长度是可变的。如果把它应用到数字传真机上进行游程长度编码时, 则规定出现概率大的游程长度用短的码字, 出现概率小的游程长度用长的码字。

霍夫曼编码的原则

(一)任何两个消息(这里指的是游程长度)其编码后的码字是不同的。

(二)码字的结构不需要用额外的指示来表明一个消息的码字从何处开始到何处终止, 而能自行分开;

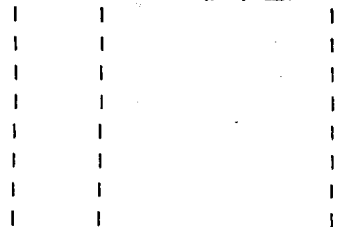
(三)出现概率大的消息用短的码字, 即:

$$P_1 > P_2 \dots \dots \dots > P_{n-1} > P_n$$

$$L_1 < L_2 \dots \dots \dots < L_{n-1} < L_n$$

式中 P_1 为 X_1 出现的概率, 也是 C_1 码字的概率。

P_2 为 X_2 出现的概率也是 C_2 码字的概率。



P_n 为 X_n 出现的概率, 也是 C_n 码字的概率。

为讨论方便, 式中消息按概率大小次序排列。 L 为码字的长度, 注脚为出现概率大小的次序。

(四)必有两个长度相同(L_n 和 L_{n-1})的消息, 它们除最后一位码字(最后一个符号)

不相同外，其余若干位码（其余符号）均应相同；

（五）长度为 L_{n-1} 的各种序列，必须用作消息码，或者必有每一序列中的一个字首用作一个消息码。

根据上述编码原则，规定霍夫曼码的编码步骤如下：

第一步，先将给定的消息按概率大小排列；

第二步，将两个最小的概率相加，其结果作为一个新的消息，按概率大小排到消息序列中；

第三步，重复第二步的步骤，一直进行到概率之和等于 1 为止；

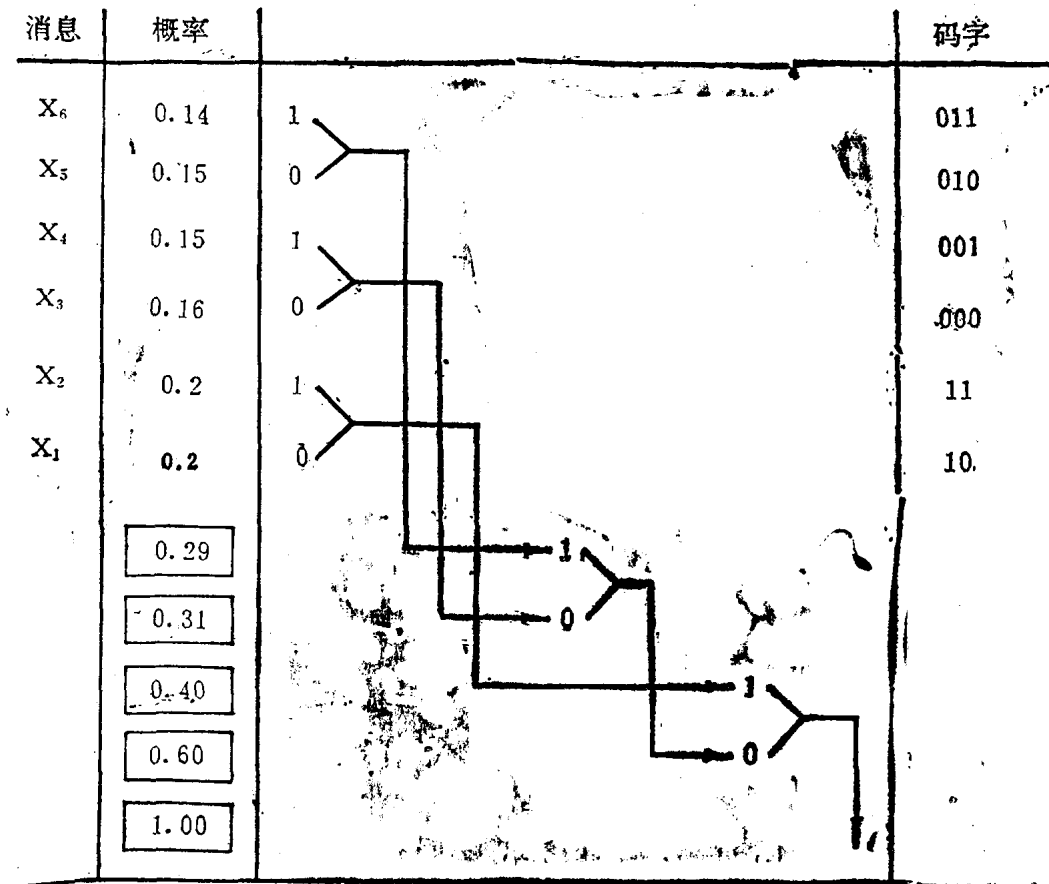
第四步，按概率顺序由小到大，每两个消息分配一组符号，一个分配“1”，另一个分配“0”。

第五步，把每个消息所分配到的符号连起来，即为霍夫曼码。

例（1）将信源 $\rightarrow \{X\}$ ， $P(X)$ ：

X_1 X_2 X_3 X_4 X_5 X_6

0.2 0.2 0.16 0.15 0.15 0.14 编成霍夫曼码，如〔图2.3.2〕所示。



〔图2.3.2〕霍夫曼编码

一种编码方案的好坏，在信息论中是用它的编码效率衡量的。码字的熵和码字的平均