高等学校試用教科书

信息论

編著者：蔡长年　汪潤生
选編者：邮电学院教材选編組信息論小組

人民邮电出版社

目　录

第二部分　信息传輸的有效性

第三部分　抗干扰性論

第四部分 信息传輸的可靠性

第一部分　通信的数学理論

第一章　离散消息和它的傳輸

§1. 引　言

信息論与很多科学技术有关系。实际上連談話、信、报紙、书籍、电影……等等都是信息传輸的形式。人們虽然不一定通晓信息論，但在生活实践中都接触到信息传輸問題。常常听到人說得到不少的信息。但是信息的多少究竟是用一种什么样的“量”来量？用什么单位？却很少人注意到这一問題。用一件、两件，用語言长短、文字的多少都不恰当。其原因有二。第一，对各式各样的消息应有統一度量的方法和单位。第二，考虑下面两种情况。如果有一个人与你約好，在指定时間准給你一个电报，电文預先你全知道，对你来說从电报里沒有得到任何信息。如果电文你不知道，当然你就得到了信息。两种情况的区别何在？在前一种情况，从概率論的观点看，对方送出来的那个电文是必然事件；而在第二种情况，电文是不定的（随机事件），并且可能性越小（事先几乎意想不到的）的电文似乎它的意义应該越大。所以，信息量与电文的概率有联系。

設发送消息的一端要把一个发生的事件告訴接收端。为了簡单起見，所发的消息只描述这个事件，而且它的描述精确地切合該事件，这样該消息就完全等价于該事件。从概率的意义上說，也就是使該消息出現的概率等于該事件出現的概率。这样的条件也符合于設計通信系統的要求，因为通信系統的任务就是要尽快、尽可靠地传送某一种类型所有可能的消息（例如对話时所有可能的話語，电視节目中所有可能的图象等等），而不管每一个消息与所要描述的事件是否切合。在上述前提下，我們将收到这个个别的消息时所收到的信息量定义为：

$$\text{接收到的信息量 } I=\log_a\frac{\left[\begin{array}{l}\text{消息传到以后接受者所視}\\\text{某一事件的事后概率}\end{array}\right]}{\left[\begin{array}{l}\text{消息传到以前接受者所視}\\\text{同一事件的先驗概率}\end{array}\right]}=\frac{P_B}{P_A}. \tag{1·1}$$

这个定义不完全恰当，暂时先不談。引入这样一个定义的理由部分地是为了陈述問題的方便。将消息所含的信息量与消息传到以后与以前某事件出現概率的变化联系起来合不合理？应該說它是符合于人們日常的概念的。这在下定义之前已經大致談过了，現在再具体地来考虑它的含意，如果接收到的消息所指的事件是你完全知道的，則式(1.1)分子及分母里所指的事件在你看都是“必然事件”，其概率均为 1，故 $I=0$；若所收到的消息几乎完全出于你意料，則分母的概率很小，分子的概率仍然为 1，所以 I 很大。現在讀者一定要問：“分子旣然一定是 1，为什么不直接了当地写成 1？”实际上分子并不一定是 1，如果在传输途径中有噪声，信号就会有畸变，畸变大到一定程度，就可能使消息发生錯誤，那么分子就不一定是 1 了。举一个例会更明白些。例如，发一个只有一个汉字的电报“7193 (电)”，在消息未传到以前，从接收端的观点看“电”字的概率是 1/10,000。如果传输途径中沒有噪声，那么传到以后是“电”字的事后概率是 1。若传输途径中有噪声，可能使其中一个数碼錯誤，則接收端收到某一电碼以后，不能肯定有无錯誤；在它看，所发送的是“电”字的概率虽然不再是万分之一，但也不是 1。例如收到的是 7103 受信者并不知道是否有錯誤，故认为所发的字可能是 7103 或是与 7103 差一数碼的 36 个字中（其中包括 7193）的一个。如果噪声足以造成两个数碼的錯誤；那么分子便更小。在整个传輸系統內均无噪声时，分子才是 1，这时，

$$I(\text{无噪声传输})=-\log_a\left[\begin{array}{l}\text{消息传到以前，它所描}\\\text{述的事件的先驗概率}\end{array}\right] \tag{1·2}$$

公式中引用对数的用意与通信中应用功率比的对数作为传輸效率相似。传輸一个汉字（假定 1,2,···0 每一数字的出現概率都是

1/10)时，

$$I\text{(无噪声)}=-\log_a\left[\frac{1}{10^4}\right]=\log_a 10^4.$$

如果传送两个汉字（暂时抛开語言中前后两个字常常有一定联系，有些字几乎永远联不到一起这样較复杂的因素不談），则

$$I\text{(无噪声)}=-\log_a\left[\frac{1}{10^4\cdot 10^4}\right]=2\log_a 10^4;$$

即 I 与消息的长短成正比，或 I 与消息的长度成綫性关系。如果应用拼音文字，则在用打字式电报机时一个字母的信息量为 $\log_a 2^5$。推广到一般情况，消息由 m 个符号組成，每一个符号可能有 L 种情况，或者說它是从一个由 L 个符号組成的符号表中抽出来的，每种符号出现的概率都是 $\frac{1}{L}$，则

$$I\text{(无噪声)}=\log_a L^m=m\log_a L \tag{1·3}$$

上面所說的一个汉字的例子中 $L=10$，$m=4$，而打字式电报中 $L=2$，一个字母的 $m=5$。应該注意，式(1·3)对于**消息集**中的任何一个消息（任一汉字或任一字母）都一样，并非針对个别消息。剩下来的問題是“对数的底 a 应該是什么？”这与信息量所用的单位有关。当然可使 $a=10$，但一般不用。现在都通用 2。因为把这样一个消息当作是最簡单的消息是合理的，即：消息只有一个单元，而这个单元只能是两个等概率的符号$\left(\text{概率各为}\frac{1}{2}\right)$中的一个。这样，如令 $a=2$，每一消息的信息量 $I=1$。单位是“二进单位”。

现在簡单地說明一下，为什么在給信息量的定义以前，先說明了一个前提——消息精确地描述了所要告訴接受者的事件。这是因为我們要将这本书的討論限于通信的技术問題，即消息如何正确地传輸問題。至于有关消息与它所描述的事件不完全等价或者接受者收到正确的消息以后能不能精确地理解等問題，则属于信息論中另一类問題——有些信息論研究者称它为“意义”問題。当涉及到这种問題时公式(1·1)中所用的事件的概率就不一定等于描述它的消息

的概率，因此，如果公式(1·1)中所用的是事件的概率，则 I 称为“意义”信息量；若用消息的概率，则称为“語言”信息量。因为我們早已訂下一个条件，消息完全等价于它所描述的事件，所以两者并无区别。从通信的技术問題看，只須研究消息如何正确地传輸，所以以后就不再区别这两种信息量的概念，也毋須在信息量前冠以“語言”二字。与此相似的情况是所传到的消息是否是接受者所关心的，或者所用符号（語言等）是否为接受者所懂得的，以及其他涉及到与接受者的主观有关的問題，都不在討論之列。因为从消息传輸的技术問題来看，这些并不須涉及。最后还要說明，当計算信息量时，在意思上完全相同，但符号（例如单字）排列不同的消息要当着不同的消息处理。例如“生产了 30 万吨鋼和 50 万吨生鉄”与“生产了50万吨生鉄和30万吨鋼”，虽然意义相同，但被认为是两个消息，各有出現概率。

§ 2. 符号(消息)平均信息量、不定度和“熵”

任何消息不論它多么长（只要不是无限长），总是从一組有限数目的消息所組成的集里选择出来的。集里的每一消息具有一定的概率。集里消息的数目可能很大，但总是有限的。例如，以汉字构成的消息，都是从所有汉字（按电报中所用的不满一万）中逐个选择排列而成，不論消息包含多少字，因为已設它的长度是有限的，所以消息的数目是有限的；尽管我們說不出消息集里所包含的消息有多少。

在上节中我們定义了一个消息的信息量，并指出这个定义有缺陷。問題在于通信系統应該准备传送任何消息，而不是針对某一个消息而設計。因此，計算信息量时应針对消息集里的所有的消息。严格地說，說“某一个消息的信息量”是不恰当的，**消息的信息量应該是指某一个范圍内的平均值而言。**

設某一信息源产生由 m 个符号組成的消息，每一个符号都是从一个由 L 个符号所組成的符号表中选择出来的，前后的选择相

互独立，即前一个符号的选择不影响以后任何一个的概率分布。当 L 个符号出現的概率相等时，每个的概率都是 $\frac{1}{L}$。前节已討論过，任一有 m 个符号組成的消息的出現概率是 $\left(\frac{1}{L}\right)^m$ 而其信息量是 $-\log\left(\frac{1}{L}\right)^m = m\log L$。因为每个符号的出現概率都是 $\frac{1}{L}$，前后的选择独立，所以在这个消息集中共有，

$$M = L^m \tag{1·4}$$

个消息，每个消息的概率都是

$$\frac{1}{L^m}$$

因此对每一个消息說信息量都一样。这样，对全消息集求平均也是同一数值；因此，可以写成

$$I = \log M. \tag{1·5}$$

每个符号的信息量是

$$I_1 = \frac{I}{m} = \log L. \tag{1·6}$$

这个概念也可以应用到由 m 个等概率的基本消息所組成的消息的情況。基本消息集內消息的数目为 L，而前后基本消息的选择也是相互独立的，这时 I 就代表每一基本消息的信息量。例如，可以把代表一个汉字的数碼看成一个基本消息，而每个基本消息由 4 个符号組成。这里基本消息集內共有10,000个消息。每个汉字的信息量是 $\log 10^4 = 4\log 10$。① 每个基本消息由 4 个符号組成，符号表內共有10个符号，每个符号的信息量是 $\log 10$。

現在进一步来考虑这样的情况。消息还是从有 L 个符号的符号表中选择 m 个来組成，前后的选择相互独立，但每个符号的概率分别为 $p_1, p_2, \cdots p_L$。我們仍然假定每个符号在时間上是等长的，

① 这里暫时假定每一数碼的出現概率相等。

这仅仅是为了使消息的长度（以符号数目度量）与延續时間成比例。設在某一消息的 m 个符号中，符号表内的第 i 个符号的数目是 m_i，那么

$$\sum_{i=1}^{L} m_i = m.$$

显然，这个由 m_1 个第一符号，m_2 个第二符号，……所組成的消息的概率是

$$p = \prod_{i=1}^{L} p_i^{m_i} \tag{1·7}$$

当任何符号的数目变动时，消息的概率就不一样。在 m 不大时，每个符号在消息中各出現几次是很难說的。因此，要針对消息集求一个統一的符号信息量是困难的。在这样的情况下，須設 m 充份地大，严格地說，使 $m\to\infty$ 时，来求符号信息量 I_1。当 $m\to\infty$，根据大数定律，m_i 以概率 1 取 mp_i，所以

$$p = \prod_{i=1}^{L} p_i^{mp_i} \tag{1.8}$$

并且，所有的消息在符号的成份上都相同，所謂成份就是每个符号所占的百分比；因而可能消息的总数

$$M = \frac{1}{p} = \frac{1}{\prod\limits_{i=1}^{L} p_i^{mp_i}} \tag{1·9}$$

$$I = \log\frac{1}{p} = \log M = -m\sum_{i=1}^{L} p_i \log p_i \tag{1·10}$$

$$I_1 = \frac{I}{m} = H = -\sum_{i=1}^{L} p_i \log p_i. \tag{1·11}$$

这里可以看出，**可能消息的总数**

$$M = 2^{mH} = 2^{(H/t_0)T} \tag{1·9a}$$

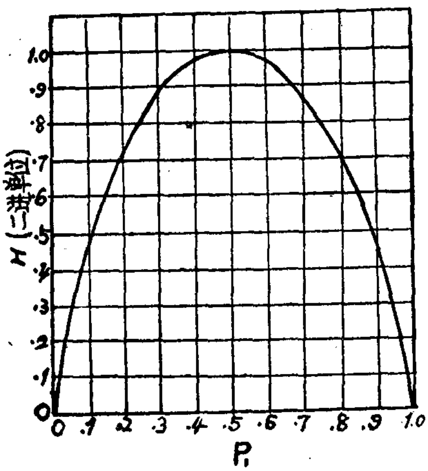
是 m 的指数函数，也是消息延續时間 T 的指数函数，上式中的 t_0 为每个符号的延續时間。

I_1 是每个符号的平均信息量。当 m 不接近无穷大时，m_i 就不一定等于 mp_i，不同消息的概率不同，所含信息量也不一样。它将在平均值上下波动。在通信系統中，被传輸的消息一个接一个进入系統，所以对通信系統說，消息可以认为是很长的。例如，对一个电报通路来說，虽然每个电报都不一定很长，但接起来就很长；从传輸的观点看，所有电报接起来可以被看成是一个消息。因此，前面設 $m\to\infty$ 是合理的，也是符合实践的要求的。

式(1·11)所表出的 H 是諸 $p_i(i=1,2,\cdots L)$ 的函数，所以写完全了应該是

$$H(p_1,p_2,\cdots,p_n),\tag{1·12}$$

H称为"熵"。这是从热力学中借用的名詞。在这里它是信息源产生信息的速率的度量，可以称为信息源的信息率。"熵"在信息論中有重要的地位。它是信息量的度量；同时它也可作为"选择度"或所謂"不定度"的度量，这个概念只要看一下 $L=2$ 时的 $H(p_1,p_2)$ 随 p 的变化就明白（图1·1）。



$(p_1+p_2)=1$

图 1·1

当 $p_1=0$，$p_2=1$，或 $p_1=1$，$p_2=0$ 时，$H=0$。实际上，当$p_1=0$ 时，消息只能选择第二个符号；$p_2=0$ 时；消息只能选择第一个符号，都說明沒有选择度或沒有不定度。当 $p_1=p_2=1/2$时，$H=1=$最大，說明这时选择对任一符号都沒有任何傾向；这就是选择度最大，也是不定度最大的情形。由此可見，"熵"是选择

度或不定度的度量。如果推广到 L 为任何数时，則在其中一个 $p_i=1$，而其余均为 0 时，$H=0$； 当 $p_1=p_2=\cdots=p_L=1/L$ 时，H 为最大，等于$\log L$。实际上，可以令

$$p_L=1-\sum_{i=1}^{L-1}p_i,$$

$$H=-\sum_{i=1}^{L-1}p_i\log p_i-\left(1-\sum_{i=1}^{L-1}p_i\right)\log\left(1-\sum_{i=1}^{L-1}p_i\right)$$

将上式对前 $L-1$ 个 p_i 求微分，并令这些偏微分等于 0，即可找出諸 p_i 間的关系。这里将对数的底換为 e 比較便利，但是在令諸偏导数为 0 后，由于换底而使上式右端所要乘的常数因子不起作用，所以

$$-\ln p_i-1+\ln\left[1-\sum_{i=1}^{L-1}p_i\right]+1=0,\ (i=1,2,\cdots L-1)\cdot$$

由此
$$p_i=1-\sum_{i=1}^{L-1}p_i=p_L,$$

即
$$p_1=p_2=\cdots=p_L=\frac{1}{L}.$$

这就証明了上面的說法（在上式成立时H不是最小値是很明显的）。

"熵"H 有下列性质：

(1) $H\geqslant 0$;

(2) H 对 p_i 連續；

(3) 当所有 $p_i=1/L$ 时，H 是 L 的非降函数；

(4) 当选择可以划分为兩步时，則"熵" H 为两步的"熵"的加权和。

最后一个性质的意义可以通过图 1·2 理解。在左边有三个选择，"熵"是 $H\left(\frac{1}{2},\ \frac{1}{3},\ \frac{1}{6},\ \right)$；在右边，第一步有两个选择，

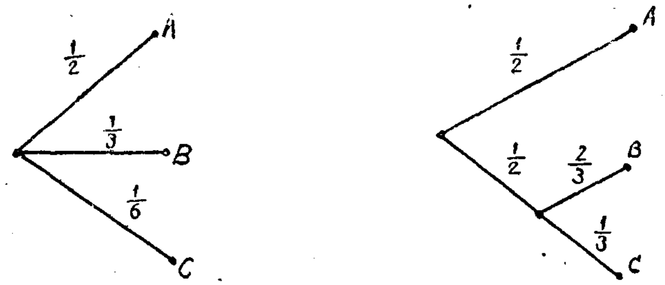


图 1·2

"熵"是 $H\left(\frac{1}{2}, \frac{1}{2}\right)$，要选择到 A、B 或 C 这一选择是必需的，所以认为加权时的权是 1。第二步在选择了非 A（即 A 的对立事件 $\bar{A}$）以后，选 B 和 C 的概率分别是 $P(B|\bar{A})=\frac{2}{3}$和 $P(C|\bar{A})=\frac{1}{3}$，所以第二步的"熵"是 $H\left(\frac{2}{3}, \frac{1}{3}\right)$。但是因为遇到这一步的概率只有 $\frac{1}{2}$，故权应为 1/2，这样

$$H\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right)=H\left(\frac{1}{2}, \frac{1}{2}\right)+\frac{1}{2}H\left(\frac{2}{3}, \frac{1}{3}\right).$$

如果把这一包含三个选择的例子写成一般化公式：

$$H(p_1,p_2,p_3)=H(p_1,p_2+p_3)+(p_2+p_3)\cdot$$
$$\cdot H\left(\frac{p_2}{p_2+p_3}, \frac{p_3}{p_2+p_3}\right).$$

这个公式可以扩充。一方面可以扩充到第一步的两个选择都可以有第二步选择；另一方面也可以使每一步的选择都不止两个可能性。

式（1·11）也可以应用于消息由若干基本消息組成的情况。这时，設可能的基本消息的数目为 S 个，相应的出現概率为 $p_1,p_2,\cdots p_s$，則每个基本消息的平均信息量，或基本消息"熵"是

$$H=-\sum_{i=1}^{s} p_i \log p_i. \tag{1·13}$$

如果每个基本消息由 q 个符号組成，則每个符号的平均信息量是

$$h=-\frac{1}{q}\sum_{i=1}^{s}p_i\log p_i. \tag{1·14}$$

例如，每个汉字的平均信息量是

$$H=-\sum_{i}p_i\log p_i, \tag{1·15}$$

求和的范围遍及所有的汉字，組成每个汉字的电碼的每个数字的平均信息量是

$$h=-\frac{1}{4}\sum_{i}p_i\log p_i. \tag{1·16}$$

§ 3. 具有固定制約的編碼制度

前面已經讲过，当符号长度相等时，所可能組成的不同消息的数目是消息长度（符号数 m 或时間 T）的指数函数。本节首先要考虑的問題是，当符号长度不等时，在一定时間 T 内可能編成的不同消息的数目 $N(T)$。当符号长度相等时，消息长度按符号数或按时間計沒有什么差别。令消息的时間长度是 T，它是从有 L 个符号的符号表中选择若干个組成的，这些符号的长度（时間上）不等，现在要在 $T\gg t_{\max}$（$t_{\max}$ 是最长符号的时間长度）的情况下来求 $N(T)$。令符号表为

符 号 名 称：$s_1\ s_2\ \cdots\ s_L$

(时間)长度：$t_1\ \ t_2\ \cdots\ t_L$

因为每一个消息的最后一个符号总是 $s_1, s_2, \cdots s_L$ 中的一个，沒有其他情况，所以可写出方程式

$$N(T)=N(T-t_1)+N(T-t_2)+\cdots+N(T-t_L); \tag{1·17}$$

但是有一附带說明。当用宗数递增的方法来計算 $N(T)$ 时，若碰到等式右端中的任一項 $N(T-t_i)=N(T_i)$ 等于 0，則必須区别两种情况。第一是宗数 $T_i=T-t_i\geqslant 0$，則須令这一項 $N(T-t_i)=1$，因为这时 $T-t_i$ 虽然不够 t_i 那么长，但是我們考虑的是长度为 T 而最后符号为 s_i 的消息；即使 $T-t_i=0$，T 也等于 t_i，所以可排出一个消息，它有一个符号 s_i。第二是 $T-t_i<0$，那么这一个 $N(T-$

$-t_i)=N(T_i)$ 一定为 0，因为 $T<t_i$，不够 s_i 的长度。具体地說，就是当（1·17）式右端的某一项为 0 时，如其形式为 $N(0)$、$N(1)$ 等，但宗数比最短符号的长度还小，須令它为 1；如系 $N(-1)$、$N(-2)$，則眞正是 0。若宗数达到最短符号长度或更大，这时 $N(T-t_i)$ 已不等于 0，故按它的应有的值計算。为了清楚起見，举一个簡单的例子，字母表是：

拼音符号：	t	n	d
长　　度：	3	5	7；

那么，

$N(T)$ 的值是：	相应的序列是：
$N(0)=0$,	无；
$N(1)=N(-2)+N(-4)+N(-6)=0$,	无；
$N(2)=N(-1)+N(-3)+N(-5)=0$,	无；
$N(3)=N(0)+N(-2)+N(-4)=$ $=1+0+0=1$,	t；
$N(4)=N(1)+N(-1)+N(-3)=$ $=1+0+0=1$,	t；
$N(5)=N(2)+N(0)+N(-2)=$ $=1+1+0=2$,	t, n；
$N(6)=N(3)+N(1)+N(-1)=$ $=1+1+0=2$,	tt, n；
$N(7)=N(4)+N(2)+N(0)=$ $=1+1+1=3$；	tt, n, d；
$N(8)=N(5)+N(3)+N(1)=$ $=2+1+1=4$；	tt, nt, tn, d；
$N(9)=N(6)+N(4)+N(2)=$ $=2+1+1=4$；	ttt, nt, tn, d；
$N(10)=N(7)+N(5)+N(3)=$ $=3+2+1=6$；	ttt, nt, dt, tn, .