科學圖書大庫

資訊理論與編碼

編譯者　郭錦鐘

徐氏基金會出版

科學圖書大庫

資訊理論與編碼

編譯者　郭錦鐘

徐氏基金會出版

編譯者序

資訊理論發展迄今，已經有三十多年的歷史了。國外在這方面的研究著作不但數量很多，且也十分深入；反觀國內的情況，則不太熱烈。直到近幾年來，才有專家學者們將它應用到統計通信、商品價格預測等方面上；或者將它納入作業研究學的探討課題之一，而逐漸蔚成研究風氣。

本書係根據史丹福大學副教授*Norman Abramson* 所著之*Information Theory and Coding*一書編譯而成。該書所涵蓋的內容十分豐富，舉凡資訊理論的起源、發展、理論推演，乃至於在音樂作曲、遺傳學……等方面上的應用，都逐一述及。遠自新約聖經，近至最現代化的人造衛星，旁徵博引，十分精彩！個人認爲這是修讀電機工程、電信工程、電腦科學、作業研究及統計應用的人所不應錯過的一本好書。

著書本不易，編譯尤難。個人雖在這本書上下了不少功夫，但囿於學植，也許仍會有誤謬不當之處；因此，懇請方家先進們能不吝指正，期使本書益臻完善。

郭　錦　鐘　謹誌

西元一九八三年六月六日

符號的彙總與熵的表示法

G-1　圖

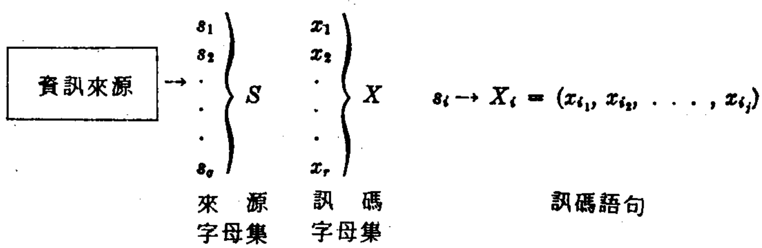

圖 G-1　編譯一個資訊來源

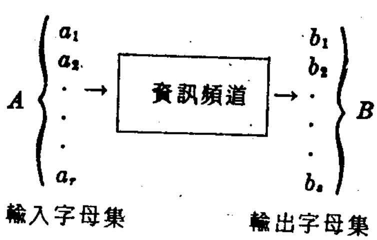

圖 G-2、一條資訊頻道

G-2　通用的符號

S　　來源字母集

s_i　　來源符號

q　　來源符號數

S^n	來源字母集爲第 n 級推廣
σ_i	來自來源之第 n 級推廣的符號
$\overline{S}$	來源 S 的伴隨
P_i	來源符號 s_i 的機率
X	訊碼字母集
x_i	訊碼符號
r	訊碼符號數（也是頻道輸入符號數）
X_i	對應於 s_i 的訊碼語句（一組 x_i 的序列）
l_i	應對於 s_i 之訊碼語句 X_i 中的訊碼符號數
λ_i	對應於 σ_i 之訊碼語句中的訊碼符號數
L	S 的一個訊碼語句之平均長度
L_n	S^n 的一個訊碼語句之平均長度
A	頻道輸入字母集
a_i	頻道輸入符號
r	頻道輸入符號數
A^n	頻道輸入字母集的第 n 級推廣
α_i	來自頻道輸入字母集之第 n 級推廣的符號
B	頻道輸出字母集
b_j	頻道輸出符號
s	頻道輸出符號數
B^n	頻道輸出字母集的第 n 級推廣
β_j	來自頻道輸出字母集之第 n 級推廣的符號
m	一個馬可夫來源的階次
P_{ij}	一個頻道矩陣元素，是 a_i 被發送而接收到 b_j 的機率
p	一條 BSC 的錯誤機率（$\bar{p}=1-p$）
C	頻道容量

P_E	錯誤機率
M	一套訊碼中的訊息數
R	資訊率
D	韓明距離
$d(b_j)$	決策規則

G-3 熵的表示法

$I(s_i) = \log \frac{1}{P(s_i)}$ 接收到 s_i 時所獲得的資訊量(無記憶來源)

$I(s_i/s_j) = \log \frac{1}{P(s_i/s_j)}$ 已經接收到 s_j 以後,再接收到 s_i 時所獲得的資訊量(第一階馬可夫來源)

$H(S) = \sum_{S} P(s_i) \log \frac{1}{P(s_i)}$ 無記憶來源 S 的熵

$H(S/s_j) = \sum_{i=1}^{q} P(s_i/s_j) \log \frac{1}{P(s_i/s_j)}$ 第一階馬可夫來源 S 的條件熵

$H(S) = \sum_{S^2} P(s_i, s_j) \log \frac{1}{P(s_i/s_j)}$ 第一階馬可夫來源 S 的熵

$H_r(S) = \frac{H(S)}{\log r}$ 以 r 進位單位衡量的熵

$H(\omega) = \omega \log \frac{1}{\omega} + (1-\omega) \log \frac{1}{1-\omega}$ 熵函數(圖 2-3)

$H(A) = \sum_{A} P(a) \log \frac{1}{P(a)}$ 輸入字母集 A 的熵(事前熵)

$H(A/b_j) = \sum_{A} P(a/b_j) \log \frac{1}{P(a/b_j)}$ A 的條件熵(事後熵)

$H(A/B) = \sum_{A,B} P(a, b) \log \frac{1}{P(a/b)}$ A 對 B 的雙關值

$H(A, B) = \sum_{A,B} P(a, b) \log \frac{1}{P(a, b)}$　　A與B的聯合熵

$I(A; B) = H(A) - H(A/B)$　　A與B的共同資訊

$I(a; B) = \sum_{B} P(b/a) \log \frac{P(b/a)}{P(b)}$　　條件共同資訊

$H(A, B/C) = \sum_{A,B,C} P(a, b, c) \log \frac{1}{P(a, b/c)}$　　A與B對C的雙關値

$H(A/B, C) = \sum_{A,B,C} P(a, b, c) \log \frac{1}{P(a/b, c)}$　　A對B與C的雙關値

$I(A; B/C) = H(A/C) - H(A/B, C)$　　已知C時，A與B的共同資訊

$I(A; B; C) = I(A; B) - I(A; B/C)$　　A，B與C的共同資訊

目　　錄

第一章 緒　論

1-1　什麼不是資訊理論

將資訊理論（*information theory*）劃歸爲一個科學的學門時，它是一個相當吸引人的名稱；而當以它作爲本書的主題時，它也多少有點令人眩惑。資訊理論的起源，得回溯到夏農（*Claude E. Shannon*）於1948年在貝爾系統技術學報（*Bell System Technical Journal*）上所發表的一篇論文（夏農，1948）*。也許夏農已經了解資訊（*information*）這個字眼之令人眩惑的本質，因此將他的論文稱作通信的數學理論（*A Mathematical theory of Communication*）。就通俗的資訊意義而言，夏農的論文所討論的乃是資訊的傳遞者—符號—而非資訊本身，它所討論的是通信與通信的方法，而非通信最終的，且令人難以捉摸的產物—資訊。

我們在此所作的這項區別是很重要的，從第2章起，我們將導出用以傳送資訊之符號的許多性質。我們將會明白：假如想要使符號能傳送資訊的話，那麼它們必須遵循某些法則。我們將使符號的性質，與它們所能傳送的資訊量發生關聯。然而，一個考慮中的符號是否眞能傳送資訊，則將視我們理論範疇以外的因素而定。例如，“*le soleil brille*”只能將資訊傳送給本書的某些讀者而已，而共同的語言則便利了資訊的傳送。能影響資訊的因素當中，較不明顯的便是心理因素。“陽光普照”這個句子若讓一位精神病患者聽到了，我們不難想像它也許隱涵了氣象方面以外的意義。語義上的因素（*semantic factors*）有可能使同一組的字對不同的聽者傳送不同的意義。夏農（1948）

* 括弧中所提到的參考書可以在本書後面的參考書目中找得到。

已經說過：通信的語義面與工程問題是無關的。然而韋佛（*Weaver*）（1949）指出：其逆敍述不一定爲眞──通信的工程（或技術）而也許與語義，心理及語言面有關。我們將在2-8節中說明本書中所發出來的理論在語言學上的應用。除了2-8節及每一章後面的某些註解以外，我們將不再探討資訊理論在其他領域上更專門性的應用。

我們將探討資訊理論的中心觀念，並特別著重於資訊的衡量法及其意義的詮釋。讀者們也許希望更詳盡地去深究資訊理論在其他領域上的應用，就這點而言，其可能性是無窮盡的。我們可以使本書中所討論的東西，與統計實驗所提供的資訊發生關聯〔林德列（*Lindley*），1956；庫爾貝克（*Kullback*），1959；葛瑞登柏格（*Grettenberg*），1962〕。我們將發現：此處所探討之資訊理論的中心觀念熵（*entropy*），至少在形式上與熱力學（*thermodynamics*）中的熵是相當的〔布里洛因（*Brillouin*），1956；賈尼斯（*Jaynes*），1959〕。資訊理論在心理學〔奎斯特勒（*Quastler*），1959〕，藝術〔皮爾斯（*Pierce*），1961，250－267頁〕及語義學〔巴─希列爾與卡納普（*Bar-Hillel and Carnap*），1952〕上的應用也有人研究過。最後，讀者們可以去參閱資訊理論在神學方面刺激而有趣的討論〔伊里爾斯（*Elias*），1958〕。

1-2 什麼是資訊理論

我們研究資訊的第一步，將是定義一種資訊的衡量法，並去探討這種衡量法的性質。我們所探討的性質將會使我們的衡量法顯得很合理，並有助於使數學理論與激發這理論的實體模式發生關聯。然而請注意：對於我們資訊衡量法之定義的認可（*justification*），是無法從完全包含於整個定義架構中的關係而獲得的。很明顯地，我們已經建立了一個本身一貫而合理的資訊理論架構；然而這個未經進一步認可的架構，充其量也只不過是構成了一門數理的學科而已。我們只有在所建立的架構和與此架構完全無關之數量的關係中，才能獲得對於這理論的認可。因此，我們將導出一種資訊的定義，並爲這個本身十分合理的定義導出一組性質來。然而，資訊的定義是無法由關係的內在一致性而加以認可

的，而須藉著說明這些關係如何應用到不包含於資訊理論架構中的數量上，才能加以認可。為了強調我們的數學模式（*mathematical model*）與眞實世界間之對應關係的必要性，我們將利用這一章緒論性的東西來質問幾個其生成與任何特定之資訊衡量法無關的重要問題。在第2、3及4章中，我們將會看到：我們對於資訊所下的定義，是如何地對這些問題提供了數量化的，且在數理上極為漂亮的解答。

1-3　編譯資訊

讓我們來考慮一些資訊傳送的例子，以便介紹資訊理論的基本概念。首先，我們將侷限於考慮一種特別簡單，但却很重要的資訊──二元資訊（*binary information*）。二元資訊的例子很容易徵引，儲存在打卡孔片（*punched cards*）中的資訊，開關式電傳打字系統（*on-or-off teletype system*）所傳送的訊息，以及儲存於電腦中之雙穩態元件（*bistable element*）內的資訊，都只不過是一些這一類的例子而已。若將本章其餘的部份都侷限在這種形式的資訊上，則我們將可以大大地簡化了我們所要闡明的論點。

很有趣的是，請注意到：和一般人的想法恰恰相反，資訊的二元表示法並非十分新穎的，而是長久以來人們所熟知的了。事實上，早期提到二元資訊之重要性的文字，乃是出現在聖經馬太福音第五章第三十七節中：如果是，你們就說是；如果不是，你們就說不是：若再多說，便是出於那惡者”（“*But let your communication be Yea, yea; Nay, nay: for whatsoever is more than these cometh of evil*”）。這種觀點也許是太極端了一點，但是從第2章起，我們將要就二元資訊與非二元資訊（*nonbinary information*）來考慮資訊理論。以二進位數元（*binary digits*）0與1表示非二元資訊的一個簡單的例子顯示在表1-1中。

表1-1中所示之二元序列與十進位數元（*decimal digits*）之間的對應關係，便是一套訊碼（*code*）的一個簡單例子。表1-1中那10組二元序列稱為訊碼語句（*code words*），而那10個十進位數元則稱為訊息符號（*message symbols*）。我們將在3-1節中更審愼地去

表 1-1 十進位數元的二元編碼

十進位數元	二元表示法
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001

爲訊碼與訊碼語句下定義，然而在目前的討論中則不妨粗略些。很顯然地，我們可以利用表 1-1 的那套訊碼，去就任一組十進位數元（訊息）序列而求出一組二進位數元序列。反過來說，我們也可以將一組來自這套訊碼的二進位數元序列予以回推，而得出唯一的一組十進位數元序列。

將一串二元訊碼語句回推到爲對應的訊息符號，這並非永遠都是這麼直接易得的。例如，考慮由表 1-2 所定義的訊碼。

表 1-2 一套二元訊碼

訊息符號	訊碼語句
s_1	0
s_2	01
s_3	001
s_4	111

假如已知一組來自這套訊碼的訊碼語句序列，則我們也許無法回推爲唯一的一組訊息符號。二元序列

$$111001 \tag{1-1}$$

也許是來自

$$s_4s_3 \tag{1-2}$$

或者是來自

$$s_4s_1s_2 \tag{1-3}$$

在此，讀者們也許會反對而指出：只需插入一個逗號（或空格）就可以了。當然，這是事實；然而，一旦使用了逗號（或空格），則便與我們二元訊碼的假設相牴觸了。假如我們使用逗號去分隔訊碼語句，則實際上我們已經就這套訊碼使用了三個符號—0、1與逗號。

我們很容易找出一套可以避免遭遇到表1-2那套訊碼之難題的訊碼。假如已知表1-3之訊碼的一組語句序列，則我們可以加以回推，而決定唯一的一組訊息符號。在本章中，我們只考慮這種訊碼。

表1-3 一套二元訊碼

訊息符號	訊碼語句
s_1	0
s_2	10
s_3	110
s_4	1110

1-4 資訊傳送上的一個問題

爲了要說明一些編碼的觀念以及它們與資訊衡量法之間的關係起見，讓我們考慮下面的問題。我們希望建立一套舊金山（*San Francisco*）與紐約（*New York*）間的通信系統，這套系統將於預定的時距內將舊金山的天氣狀態傳送出去，而且只能使用開關式（也就是二元）設備。爲了簡化起見，我們希望將舊金山的天氣劃分爲四種可能情況—晴天、陰天、雨天或有霧（*foggy*）。我們可以考慮如表1-4中所示的，將可能情況組成四個訊息符號，同時顯示在表1-4中的是每一種狀態的假定機率。我們假設四種狀態爲均等可能的（*equiprobable*）。

表 1-4 舊金山的天氣狀態

訊息	機率
晴天	¼
陰天	¼
雨天	¼
有霧	¼

一種將這些訊息編譯成一組二元符號序列的可能方法，便是去構建下列的對應關係，稱之為訊碼 $\mathcal{A}$。

訊碼 $\mathcal{A}$

晴天……………… 00
陰天……………… 01　　　(1-4)
雨天……………… 10
有霧……………… 11

因此，利用訊碼 $\mathcal{A}$，我們便可以將"晴天，有霧，有霧，陰天"編譯為"00111101"。

很顯然地，就已知一組訊碼語句序列，而我們須加以回推而重新建立唯一的一組產生那組特定的訊碼語句序列之訊息序列這層意義而言，訊碼 $\mathcal{A}$ 是傳送這種資訊的一套可接受的訊碼。

同時也很明顯的是：若使用訊碼 $\mathcal{A}$，則我們必須就每一道訊息發送2個二進位數元〔比尼特（*binit*）〕*。此外，讀者們不難了解：我們再也找不到另一套容許我們平均每一道訊息使用少於2個比尼特之可接受的訊碼了。

現在考慮一位洛杉磯（*Los Angeles*）的工程師所面臨的一個類

*從現在起，我們將用 *binit* 這個縮寫字來代表 *binary digit*。將比尼特（*binit*）(*binary digit*）與比特（*bit*）（我們將在第二章中加以定義的一種資訊量之單位）予以區別是很重要的。正如我們將明白的，在某些情況下，一個比尼特也許包含了一比特的資訊。

似問題。也就是說，我們希望建立一套類似的二元通信系統，以便將洛杉磯的天氣狀態傳送到紐約去。我們知道：舊金山的天氣與洛杉磯的天氣之間有很重大的氣象差異，其中之一則可以藉著將洛杉磯的天氣劃分爲晴天、陰天、雨天或煙霧（*smoggy*）而加以說明。現在對於這幾個城市中任一個城市的一位居民來說，雖然有霧與煙霧之間並沒有什麽很大的差異，但這似乎不是設計一套通信系統時的一個考慮因素。從通信的觀點來看，既然那四種狀態已經被編譯成二元序列，因而任一組特定序列的重要性或意義也就無關緊要了。

然而，另一種確實與通信問題無關的氣象差異却有可能存在。爲了對洛杉磯的氣候表示一視同仁起見，我們必須對那四種可能的狀態賦予不同的機率，這已經顯示在表 1-5 中了。

表 1-5　洛杉磯的天氣狀態

訊 息	機 率
晴 天	$\frac{1}{4}$
陰 天	$\frac{1}{8}$
雨 天	$\frac{1}{8}$
煙 霧	$\frac{1}{2}$

假如我們使用訊碼 $\mathscr{A}$ 去傳送這種資訊，則我們可以表現得與那套從舊金山傳送出資訊的通信系統一樣好，但却無法比它更好。也就是說，假如我們使用了訊碼 $\mathscr{A}$，則不管天氣的狀態如何，平均每一道訊息我們都得發送 2 個比尼特。然而，試考慮使用下列稱爲訊碼 $\mathscr{B}$ 的訊碼去傳送資訊的可能性：

訊碼 $\mathscr{B}$

晴天	10	
陰天	110	(1-5)
雨天	1110	
煙霧	0	