# 資訊的數值傳遞與檢誤

國立台灣教育學院科學教育系數學組　林濟卿

資訊的範圍包羅萬象，如聲音、形象、色彩等等，這些資訊的傳遞記錄與分析實已構成了我們日常生活的一大部分。顯然資訊傳遞活動與人類史共生，並非如今才有的新穎玩意兒，只是以前的傳統方法幾乎完全依賴於人類的天然感官，故其傳遞的量與速度均不及今日的計算機所能提供的萬分之一。

現在的資訊傳遞可說有一大部分非靠計算機的幫忙很難完成，所開發出來的各種新方法也莫不以擁有計算機爲前提，如此相輔相成的構成了一幅浩瀚壯觀的資訊景象，也使資訊傳遞這樁事完全的被納進計算機的專門領域裏去。

計算機的神速處理能力突破了傳統的直接傳遞方法而代之以離散的數值傳遞方法。現實的現象可以說幾乎都是連綿不斷的，當然人類的社會活動也不能例外，因此很自然的這些資訊的初期傳遞記錄也就採取了連續的直接方法，例如唱片的聲紋，圖像的整張傳遞與保存以及聲波的傳送等，其眞實性與不確性似乎也遠勝於使用形像、文字等描述的方法。其實並不盡然，由於製作技術的限制，唱片的聲紋，整張的圖像也不見得會處處都能滿足我們的要求標準。何況經過長久時間的折磨而變成模糊不清的記錄，要使其復元縱然不是不可能也可說是極爲困難的。

反之離散的數值傳遞就沒有這個缺點。因爲它是利用相異分明的數字符號來傳遞的，例如樂譜記錄其正確性就應比唱片記錄優越，有了模糊損傷也較爲容易復元。當然完全沒有臨場感是它的先天性缺點之一，同時它（樂譜）的複雜及高度專業性也就無法使它成爲數值傳遞的主要角色。計算機所採用的是再簡單沒有的“0”或“1”，兩者截然不同毫不含糊，因此其復元性也就優異無比。但它也有其先天不可避免的缺點那就是是動不動就有一大堆的龐大資料須要處理。可是若能克服此難關並能不嫌其煩的仔細分類，例如照片的傳眞只要能分爲夠多的黑白小點，一定可以達到預先所要求的正確標準，比之整張傳遞惟有過之而無不及。

由此可知數值傳遞的關鍵就在於這些龐大的資訊量與其所可能帶來的錯誤。本來計算機的犯錯率都相當的低，但因爲傳遞數次太多了，累積起來的錯誤也就不容忽視，偏偏這種傳遞錯誤又是不可能完全避免的，只有盡量的減低而已，於是如何檢誤便成爲數值傳遞的重要課題了。

最簡單的檢誤法可說是 Parity check。如在實行4個bit的數值傳遞時多加一個檢誤 bit 0 0000, 0 1010, 1 1110, 1 0111, …等使在此5個bit中的“1”個數都成爲偶數。計算機在接受傳遞的同時也檢查每5個bit裏的“1”個數，若不是偶數便是傳遞錯誤立即發出警號要求

處理。如果犯錯次數是偶數此法便失效，不過上面已提起，計算機的信度頗高，連續發生兩次以上的錯誤機會甚低，因此雖然只是1個bit檢查1次錯誤也有其令人重視的價值。當然能夠有2次以上的檢誤能力更好，不過那就得增加檢誤bit的個數，隨之所付出的代價便是計算機可用容量的減少與處理速度的減低。此外尚有使用兩架計算機同時傳遞相同的資訊，或是來回的重覆傳遞兩次以便互相核對等等方法，但總免不了要付出相對的代價"加重容量的負擔，犧牲處理的速度"。單是錯誤的檢出就有這麼多的困難，若還想加以錯誤修正那不就更困難嗎？的確不錯，更難，可是還是有辦法。以下便介紹一種既能檢誤(error-checking)又能改誤(error-correcting)的最簡單方法Hamming code $C_{7,4}$ 法。

設 $V_n = \{(a_1 \cdots a_n)\}$

$a_i = 0$ 或 1 $\quad i = 1, \cdots, n$

這樣的集合爲有限集合其種類共有 $2^n$ 種

再規定兩種演算 $1 + 1 = 0$

$0 + 1 = 1 + 0 = 1$

則 $V_n$ 構成一個有限的向量空間，其係數體當然是 $\{0,1\}$ 。

假定我們是採用一語(7 bit)即 $V_7$ $\{(a_1 \cdots a_7)\}$ 的集合來作數值傳遞則其種類共有 $2^7 = 128$ 種，不管是正確或錯誤的傳遞祇要是一語7 bit便逃不出這種128種範圍。

我們若能再利用此128種構成滿足下面3個條件的分類便能達到既能檢誤又能修正目的的數值傳遞碼。

1.平分128種成爲互不重覆的數類，如平分成爲16類每類8種。

2.同一類裏的每一種彼此間祇相異1個bit，如(11110000)與(01110000)；…等。

3.從每一類各選出一種代表出來所組成的集合自成爲一個向量空間。

例如 從 $V_7 = \{(a_1 \cdots a_7)\}$ 集合中選出

$v_1 = (1000011)$

$v_2 = (0100101)$

$v_3 = (0010110)$

$v_4 = (0001111)$

爲最基本的4種再施予加法演算便可造成下面的16種(含(0000000))稱爲Hamming code檢識碼以 $C_{7,4}$ 表示

$v_1 = (1000011)$

$v_2 = (0100101)$

$v_3 = (0010110)$

$v_4 = (0001111)$

$v_1 + v_2 = (1100110)$

$v_1 + v_3 = (1010101)$

$v_1 + v_4 = (1001100)$

$v_2 + v_3 = (0110011)$

$v_2 + v_4 = (0101010)$

$v_3 + v_4 = (0011001)$

$v_1 + v_2 + v_3 = (1110000)$

$v_1 + v_2 + v_4 = (1101001)$

$v_1 + v_3 + v_4 = (1011010)$

$v_2 + v_3 + v_4 = (0111100)$

$v_1 + v_2 + v_3 + v_4 = (1111111)$

$0 = (0000000)$

我們很快便可驗出此16種檢誤碼構成一個向量空間，也就是 $V_7$ 的部分空間。在 $V_7$ 中與0=(0000000)只相異一個bit的有(1000000)，(0100000)，(0010000)，(0001000)，(0000100)，(0000010)，(0000001) 7種，即從7選1的組合數

$$\binom{7}{1} = \frac{7!}{6!} = 7$$

連同本身的0一共有8種。同理可知在 $V_7$ 中與其他15種 $v_1, \cdots, v_1 + v_2 + v_3 + v_4$ 等各只相異一個bit的均爲8種，總計 $16 \times 8$

$= 128 = 2^7$ 種，含蓋了 $V_7$ 全體，而且每類每種都沒有重覆（證明見後）。

就以此 16 種檢誤碼作爲數值傳遞碼。計算機每接受到一碼（7個 bit）使與此 16 種檢誤碼 $C_{7,4}$ 核對，核到了就是傳遞正確，核找不到就是傳遞有誤，再繼續核查與此誤碼只相異一個bit的是那一個檢誤碼，該碼就是原來所要傳的正確碼。

例如所接到的是（0011001）$\in C_{7,4}$ 故傳遞正確。

如果所接到的是（0001100）$\notin C_{7,4}$ 故傳遞有誤，而此誤碼與16種檢誤碼之一的（1001100）只差一個 bit 得知原來所要傳遞的正確號碼應爲（1001100）。

爲了闡明上述三個關鍵性的分類條件首先規定傳遞碼向量空間 V 的距離 d 如下：

$d(u,w)$=對應相異的 bit 數

如　$u=(1000011)$　$w=(0100101)$

則 u,w 的距離　$d(u,w)=4$

如此規定的距離可滿足下面的距離公理：

1. $d(u,u)=0$
2. $d(u,w)=d(w,u)$
3. $d(u,w)\leq d(u,v)+d(v,w)$

1，2 甚爲顯然無須再說明，3 可論證如下：

設　$u=(a_1\cdots a_n)$　$w=(b_1\cdots b_n)$

$v=(c_1\cdots c_n)$

如果　$u=w=v$

則　$0=d(u,w)\leq d(u,v)+d(v,w)$ 成立

如果　u,w,v　互不相等則只針對其第一要素 $a_1$, $b_1$, $c_1$ 推論即可餘可類推，故可假定

$$a_i=b_i=c_i \qquad 2\leq i\leq n$$

也不失其一般性

i　若 $a_1=b_1\neq c_1$

則 $0=d(u,w)\leq d(u,v)+d(v,w)=2$ 成立

ii　若 $a_1\neq b_1$

則 $1=d(u,w)\leq d(u,v)+d(v,w)=1$ 成立

進而更可推出下面的關係式

定理1　$d(u,w)=d(0,u+w)$

因爲 $u+w$ 只有對應要素相異時其和才能爲 1，故得證。

也就是說要求 $d(u,w)$ 只求 u 與 w 之和的“1”個數便可，若要求 $d(0,u)$ 或 $d(0,w)$ 更簡單只算其所含的“1”個數便可。

如　$u+w=(1001101)+(1110001)$

$=(0111100)$

其和含有 4 個 1　　$\therefore d(u,w)=4$

而　u,w 各含有 4 個 1

$\therefore$　$d(0,u)=d(0,w)=4$

在 n 次元的 $V_n$ 部分空間中

$$b=\min\{d(0,v)\mid \forall v\in V_n\}$$

稱爲 $V_n$ 的最短語距

例如在部分空間

$V_4=\{(0000),(1010),(0101),(1111)\}$ 中

$d(0,v)=2,3,4$　$v\in V_4$，$v\neq 0$

我們稱（1010），（0101）爲 $V_4$ 的最短距語，而其最短語距爲 2。有了距離便可規定球，再設以〔$d(u,x)\leq r$〕表示以 u 爲球心半徑爲 r 的球內範圍的向量個數則有下面一個定理。

定理2　〔$d(u,x)\leq r$〕=〔$d(0,y)\leq r$〕

也就是說以 u 爲球心，r 爲半徑的球所含的向量個數等於以 0 爲球心，r 爲半徑的球所含的向量個數。

依定理 1　$d(u,x)=d(0,u+x)$

令　$y=u+x$　則　$d(u,x)=d(0,y)$

故得證

定理3　設部分空間 $V_n$ 中的最短語距爲 S，而

$$r=(S-1)/2$$

則以 r 爲半徑的此諸球把 $V_n$ 區分爲互不相疊的部分

可推證如下：

設有以 u，w 爲球心，半徑爲 r 的兩球互疊，而 v 爲此重疊部分裏的任一向量

則依假設 $d(u,v) \leq r$ $d(w,v) \leq r$

但 $d(u,w) \leq d(u,v) + d(w,v) \leq 2r$

$= S - 1$

$\therefore$ $d(u,w) = d(0,u+w) \leq S - 1$

此與 S 爲 $V_n$ 的最短語距矛盾，故以 u，w 爲心 $r = (S-1)/2$ 的兩球不得有相疊部份。

上面已提起 $C_{7,4}$ 的檢誤碼共有 16 個自成一個向量空間。在 $V_7$ 中與 0 = 〔 0000000 〕只差一個 bit 的向量有 7 個，再加上成爲球心的檢誤碼 0 本身構成爲 8 個向量一組。

依定理 2 與檢誤碼 $v_1$ = ( 1000011 ) 只差 1 個 bit 的也有 ( 0000011 )，( 1100011 )，…，( 1000010 ) 7 個向量，再加上成爲球心的檢誤碼 $v_1$ 本身構成另一組。

同理以另 14 個檢誤碼爲球心也可構成每組各含 8 個向量的另外 14 個組。

但這 16 組是否都互相分離而沒有重疊部分？答案是肯定的，因爲上提 $C_{7,4}$ 的 16 個檢誤碼與 0 的語距依次爲：

3,3,3,4,4,4,3,4,3,3,3,4,4,4,7,0

最短語距爲 3，依定理 3，以 $(3-1)/2=1$ 爲半徑的球把 $V_7$ 區分爲互不相疊的部分。

$16 \times 8 = 128 = 2^7$

可知此含有 8 個向量的 16 組正好等於 $V_7$ 的全部，因此任何一個 $V_7$ 的向量必屬於此 16 組中的僅一組且與成爲球心的檢誤碼只差一個 bit，於是自動檢誤自動修改的目的便在此達成。

一語 7 個 bit 本可傳遞 128 種資訊，但在 $C_{7,4}$ 中却只有 16 種檢誤碼眞的在擔任傳遞資訊的角色，也就是說用在傳遞資訊的只有 4 個 bit，還有不作傳遞資訊的 3 個 bit 可以說就是完成自動檢誤修正的代價。

不過 16 種也可以作很多事了，如彩色傳遞若採取 3 語 1 組的方法，第一語只有 2 個 bit 便可表示 3 種原色，第 2、3 語各動用 3 個 bit 便可把明暗、濃淡各分爲 5 等級而有餘，這樣一來就有 $3 \times 5^2 = 75$ 種顏色可資區別應該是夠用的了。尙有第 1，2，3 語的第 2，1，1 bit 沒用到可留爲備用。當然還是能用一語表完所有的資訊最爲理想，那就得另外想辦法尋求含有比 7 個 bit 更多 bit 的檢誤碼了。

此外 $C_{7,4}$ 的自動檢誤修正不是以一語一次僅錯一個 bit 爲前提的，如有 2 個 bit 以上的錯誤便失效而檢查不出來。不過我們可以提高計算機的精度來彌補這個短點。如計算機的誤率爲 $10^{-4}$，則連續錯誤 2 次的機率便降至 $10^{-8}$，不能不謂已是相當的微小，但不能否認這仍是一個先天性的缺點。若要能自動檢誤修正 2 個 bit 以上的錯誤就得再增加檢誤碼的 bit 數。

如 Golay 的檢誤碼 $C_{23,12}$ 便有 12 個 bit 可作爲傳遞資訊之用。其檢誤碼的最短語距爲 7 故可用 $(7-1)/2 = 3$ 爲半徑的球來把 $V_{23}$ 區分爲含有 $2^{11} = 2048$ 種向量互不相疊的 $2^{12} = 4056$ 組，可自動檢誤修正多至 3 個 bit 的錯誤，不過所費的代價也有 11 個 bit 之多。

以上所述的檢誤碼已是在實際上廣爲利用的定型手法，也是一個抽象的數學在具體的實務上發揮威力的好例子。在 $C_{7,4}$ 與 $C_{23,12}$ 間尙有一段空隙，是否也有其他同樣的巧妙分類存在，$C_{23,12}$ 以上呢？是否實用？都是一些頗堪玩味的有趣問題。我們更應上進一層努力探討比這些檢誤碼更簡便更有效的新方法才是。 □

參考書目

1 Digital computer fundamentals, Thomas C. Bartee, 4th ed, 1977.

2 Linear algebra with applications, Jeanne Agnew, Robert C. Knapp, 1979.