从本质看海明码——海明码的由来

陈冠斌

1.奇偶校验以及奇偶校验错误率

奇偶校验：每一位(包括校验位)都进行异或运算，结果为0

如x1~x7,y0 (数据位为x1~x7，校验位为y0)，则

x1 xor x2 xor ... xor x7 xor y0 = 0。

错误概率：

设共有n位，每一位的错误概率均为x

则错误k位的概率：C(n,k) \* xk \* (1-x)n-k

总概率：V = C(n,1) \* x1 \* (1-x)n-1 + C(n,3) \* x3 \* (1-x)n-3+ ...

而：

S = C(n,0) \* x0 \* (1-x)n + C(n,1) \* x1 \* (1-x)n-1 + ... + C(n,n) \* xn \* (1-x)0 = [x+(1-x)]n = 1

T = C(n,0) \* x0 \* (1-x)n - C(n,1) \* x1 \* (1-x)n-1 + C(n,2) \* x2 \* (1-x)n-2 - ... =[(1-x)-x]n = (1-2x)n

V = (S - T) / 2 = [ 1 - (1-2x)n ] /2 = x - 2\*C(n,1)\*x2 + 4\*C(n,2)\*x3 - ...

由于x较小，x后面的项大致可以忽略，值约为x。

这与C(n,1) \* x1 \* (1-x)n-1 把1-x估计为1时的值时一致的。

如x=0.00001，x2过于小了，所以一般传输错误的话，只考虑一位错误。

2. 设计可以检验错误位数的方法(海明码的由来)

若想知道错误的位置，则通过多个式子共同判断，假设数据位为x1~xn-1,全体域为F，错误位置为xc,第i(0<=i<m-1)个式子的数据位集合为Si，数据位进行异或运算，结果为yi，有：

xor {S0} = y0

xor {S1} = y1

……

xor {Sm-1} = ym-1

这些式子和方程式不一样，它比方程式多了一个约束条件，这个约束条件是x1~xn有且仅有一个值为1，其它值为0。

对于值为1的第k个式子，等价为xc∈Sk，存在错误位；

对于值为0的第k个式子，等价为xc∈‾Sk。

m个式子中，每个式子可以把全体域分为两部分，Sk和‾Sk(即F-Sk)，根据不同的yk值，选择其中一个集合。

错误位属于集合S0(‾S0) ∩ S1(‾S1) ∩ …∩ Sm-1(‾Sm-1)。

……

……

S0

‾S0

S1

‾S1

‾Sm-1

Sm-1

如：x1,x2,x3为数据位，y0,y1为检验位。

x1 xor x3 = y0 = 0 (i)

x2 xor x3 = y1 = 1 (ii)

即S0={x1,x3},y0=0 ; S1={x2,x3},y1=1。

错误位属于集合{x2}∩{x2,x3}={x2}，即错误位为x2。

设n位二进制数ym-1ym-2…y0为Y。对于任意一个数据位，它要么在Sk中，标记yk=1；要么在‾Sk中，标记yk=0。对于数据位x1~xm-1的任意一个数，若它表示m位二进制数zm-1z­m-2…z0,则第k个式子对应的值yk为zk。其中数据位x1~xn-1对应的ym-1ym-2…y0的值都不相等，共有2m­-1种可能性，其中若ym-1ym-2…y0的值都为0，则代表无错误位。

以下是n=3时的情况：

(4) x4 xor x5 xor x6 xor x7 = y2

(2) x2 xor x3 xor x6 xor x7 = y1

(1) x1 xor x3 xor x5 xor x7 = y0

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| y2 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| y1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| y0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| Y | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Error | None | x1 | x2 | x3 | x4 | x5 | x6 | x7 |

对于上述方式，ym-1,ym-2,…,y0的位置还没确定下来。可以把ym-1,ym-2,…,y0放在第2m-1,2m-2,…,20位，所有yk所在位置的值加起来，就是Y的值,即XY数据位发生错误(如果XY=0，则没有数据位发生错误)。其中第2m-1,2m-2,…,20位有且只有一次参与了运算，把它们当作校验位，可认为x2^(m-1),x2^(m-2),…,x0­的值为0。把x1,x2,x4…用x1+2^n,x2+2^n,x4+2^n…替代。如上文：

数据位为x3,x5,x6,x7,x9,x10,x11,x12。

(4) x12 xor x5 xor x6 xor x7 = y2

(2) x10 xor x11 xor x6 xor x7 = y1

(1) x9 xor x3 xor x5 xor x7 = y0

此时的这个方式称为海明码校验。

针对计算，比较有规律，可以通过硬件电路提高速度。

数据位：

从1开始，检验1位，跳转1位；[1,3,5,7,9,11……]

从2开始，检验2位，跳转2位；[2,3,6,7,10,11,……]

从4开始，检验4位，跳转4位；[4,5,6,7,12,13,14,15,……]

……

检验位：

每次乘以2，即二进制数每次左移一位。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 方式 | 检验码 | 检验位数 | 总位数 | 参与运算总次数 |
| 海明码校验 | n | 2n -1 | 2n -1+n | n\*2n-1 |