CODIGOS QUE DETECTAN Y CORRIGEN ERRORES

Por P. Gil Alvarez (*)

1. INTRODUCCION.-

En la primera parte de esta exposición, trataremos únicamente problemas para códigos binarios. Supongamos, pues, que a través de un canal de información, se transmite una secuencia de dígitos binarios. En ausencia de perturbaciones, cada dígito emitido será recibido claramente, sin error. Pero si, por algún motivo, la señal transmitida sufre deformación, un cero podrá ser recibido como uno, y un uno podrá ser recibido como cero, produciéndose así un error en la recepción que llevará a una interpretación incorrecta de la palabra recibida.

Con el fín de evitar estas situaciones que pueden a veces acarrear graves consecuencias, se introducen métodos de codificación que permitan detectar y corregir cualquier tipo de error. La idea fundamental de estos códigos es muy simple: se desea trnasmitir un conjunto de K dígitos; a éstos se añaden r nuevos dígitos (que serán llamados "de control") convenientemente elegidos, y se transmite la secuencia completa de n = K + r dígitos. El problema de la codificación consistirá entonces en la elección adecuada de estos r dígitos de control. Por supuesto, el método lleva consigo un desprecio de cierto número de palabras que podrían, en el caso de transmisión sin ruido, formar parte del código. Más precisamente; el número de posibles palabras de longitud n sería 2^n, de las

---

(*) Trabajo realizado con beca de ayuda a la investigación del Fondo IBM (Equipo formado por P. Gil Alvarez, F. J. Girón González-Torre, P. Ibarrola Muñoz, M. Sánchez García, R. Vélez Ibarrola; del Departamento de Estadística Matemática de la Facultad de Ciencias. Madrid).

cuales solamente 2^k serán utilizadas. El cociente K/n = R , denominado "tasa de información" , da una idea de la utilidad efectiva del código.

Como dos ejemplos de códigos de este tipo, señalemos:

1.a) Códigos de repetición.-

Para un conjunto de dos mensajes, se emplean como palabras iniciales del código las dadas por un sólo dígito (0,1); éste será el dígito de información, y se añaden a cada uno de ellos un número arbitrario r de dígitos de control, todos ellos iguales al dígito de información correspondiente; así, por ejemplo, para r = 4, las palabras del código serán

0 0 0 0 0 para el primer mensaje y

1 1 1 1 1 para el segundo mensaje.

En la recepción de palabras, puede aparecer cualquiera de las 2^5 secuencias binarias posibles; el sistema de traducción consistirá en traducir como 0 aquellas secuencias con mayor número de ceros que de unos, y en traducir 1 las que cumplan la condición complementaria. Surgen en este esquema, dos problemas fundamentales:

I) En realidad no suprimimos la posibilidad de error; únicamente disminuímos la probabilidad de que ésto ocurra (se supone siempre que la probabilidad de recepción incorrecta de un dígito es inferior a la probabilidad de recepción correcta). Estamos utilizando un esquema de traducción del tipo denominado "observador ideal".

II) En ocasiones (p.e. cuando r = 5 o cualquier número impar) es imposible la traducción porque la probabilidad de error en una palabra recibida (0 1 0 1 0 1) es idéntica tanto si se ha emitido 0 como si se ha emitido 1. En este caso diremos que el procedimiento de traducción es incompleto.

1.b) Códigos de control de paridad simple.

Para el código formado por las secuencias binarias de longitud K, podemos pensar en establecer un nuevo código dado de la forma siguiente: a cada palabra se le añade un dígito de control (r=1) de tal modo que la suma módulo 2 de los K+1 dígitos de la nueva palabra, sea, por ejemplo, el cero. Si en la palabra recibida, la suma de dígitos es uno (mod.2) algún error ha ocurrido en la transmisión. (El cálculo de la suma es lo que denominaremos "control de paridad") Así, en el esquema de traducción de observador ideal, toda secuencia recibida será traducible por la palabra correspondiente si el resultado del control es 0, y será intraducible si el resultado del control es 1 (detectamos el error, pero no podemos corregirlo). Se trata, pues, de un esquema incompleto, en el que se observa aún más claramente que en el ejemplo anterior, la diferencia entre ambos problemas: detección y corrección del error.

2.- DISTANCIA DE HAMMING.

2.1.-Definición.-

Sean $v_1, v_2, \ldots, v_{2^n}$ las posibles secuencias de longitud n (Puesto que cada palabra tiene n dígitos podemos representarla en forma vectorial). Sean $w_1, w_2, \ldots w_s$ las palabras incluídas en el código considerado.

Llamaremos distancia de Hamming entre las secuencias v_1 y v_2 al número de dígitos en que se diferencian(Puede comprobarse fácilmente que se cumplen las propiedades de una métrica).

Por ejemplo, si

$$v_1 = 0\,1\,0\,0\,1\,1 \quad y \quad v_2 = 1\,1\,0\,1\,1\,0$$

se tendrá:

$$d(v_1, v_2) = 3$$

Si w es la palabra emitida y v es la secuencia recibida, ha habido exactamente d(v,w) errores en la transmisión.

2.2.- Principio de distancia mínima.-

El observador ideal es aquél que traduce con el criterio de probabilidad de error mínima, es decir, para la secuencia recibida v, se traduce v por aquella palabra w tal que hace máximo el valor

$$Pr(w/v)$$

Además, como estamos considerando un canal binario simétrico (la probabilidad de recepción correcta y de error es la misma para ambos dígitos) en el cual la probabilidad de emisión de cada palabra es la misma para todas ellas, maximizar el valor anterior equivale (ASH "Information Theory"-pag.62) a maximizar la probabilidad

$$Pr(v/w)$$

Vamos a probar que el observador ideal coincide con el esquema de traducción consistente en interpretar cada secuencia recibida como la palabra del código más próxima a ella (entendiendo esta proximidad como la derivada de la distancia de Hamming). Sea α la probabilidad de que un dígito se reciba incorrectamente, $1-\alpha$ la probabilidad de recepción correcta; se tiene

$$Pr(v/w) = \alpha^{d(v,w)} (1-\alpha)^{n-d(v,w)} \qquad (2.2.1)$$

Si escribimos $d_i = d(v,w_i)$, y comparamos las probabilidades de recepción de v si se han transmitido las palabras w_1 ó w_2, resulta:

$$\frac{Pr(v/w_1)}{Pr(v/w_2)} = \frac{\alpha^{d_1}\cdot(1-\alpha)^{n-d_1}}{\alpha^{d_2}(1-\alpha)^{n-d_2}} = \left(\frac{1-\alpha}{\alpha}\right)^{d_2-d_1} \qquad (2.2.2)$$

Puesto que $1-\alpha > \alpha$ (probabilidad de error inferior a $\frac{1}{2}$), el cociente $\frac{1-\alpha}{\alpha}$ es superior a 1. Por tanto será

$$Pr(v/w_1) > Pr(v/w_2) \text{ si y sólo si es } d_1 < d_2 \qquad (2.2.3)$$

c.q.d.

3.- ERRORES MULTIPLES.- ESFERAS DE HAMMING.

Si cada secuencia recibida se traduce por la palabra más próxima a ella, es natural pensar que un buen código será aquel en que las palabras estén suficientemente "separadas". De acuerdo con esta idea, si asociamos a cada palabra una esfera (según la distancia de Hamming) formada por las secuencias que distan de ella un número menor o igual que e de unidades, el esquema de traducción consistirá en interpretar cada secuencia como la palabra que es centro de la esfera en que se encuentra, supuesto que todas las esferas son disjuntas, y en no traducirla si no pertenece a ninguna esfera.

Más precisamente, formulamos el siguiente teorema.

3.1.- Teorema

Sean $w_1, w_2, \ldots, w_s$ palabras binarias de longitud n. La condición necesaria y suficiente para que se verifique

$$d(w_i,w_j) \geq 2e+1 \quad , \quad \forall\, i \neq j \qquad (3.1.1)$$

para algún número entero positivo e, es que se puedan corregir todos los errores simples, dobles, etc. hasta los de tamaño e inclusive.

Si la condición anterior se sustituye por

$$d(w_i,w_j) \geq 2e \quad , \quad \forall\, i \neq j \qquad (3.1.2)$$

solamente son corregibles los errores de tamaño inferior o igual a e-1. Los errores de tamaño e pueden detectarse, pero no corregirse.

En efecto:

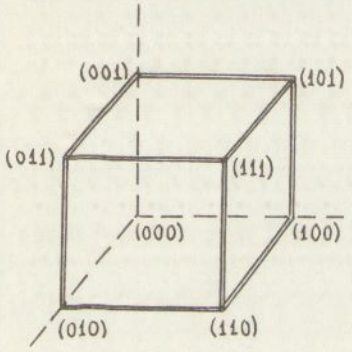
Supongamos que la mínima distancia entre palabras es 2e+1. Para que una palabra w sea traducida como otra distinta w', la secuencia recibida debe ser al menos tan próxima a w' como a w; para esto es preciso que al menos e+1 dígitos de w hayan sido erróneamente recibidos. Luego todos los errores de e o menos dígitos son corregibles (Las esferas de radio e son disjuntas).

Si la distancia entre dos palabras W y w' es 2e, un error de tamaño e en w produce una secuencia v cuya distancia a w es e, y cuya distancia a w' es al menos e, en virtud de la desigualdad triangular. Si la distancia a w' es exactamente e, v puede proceder tanto de w como de w' por medio de un error de tamaño e. Por ello, aunque el error consiga ser detectado, no es posible corregirlo.(Las esferas son, en este caso, tangentes).

Los recíprocos se siguen por un razonamiento análogo.

3.2.- Representación geométrica.-

Sea n=3. Las posibles secuencias binarias pueden representarse en los vértices de un cubo (en una dimensión sería un segmento, en dos un cuadrado y en más de tres un hipercubo).



La distancia entre dos secuencias coincide con el número de aristas que es preciso recorrer para ir de un vértice a otro.

Por ejemplo, si se toma e=1, para corregir un error simple es preciso, según el teorema anterior, una distancia mínima de tres unidades entre las palabras. Solamente hay dos palabras en el código (000 y 111, p.e.). Si se desea detectar el error simple, aun sin corregirlo, habrá cuatro palabras posibles: 100, 010, 001 y 111, o cualquier otro conjunto similar.

Se observa que al aumentar la capacidad de corrección disminuye el número de posibles palabras de código. ¿Cuál es el número máximo de palabras que puede tener un código para corregir errores de un tamaño dado?

3.3.- Cota superior de Hamming.-

Si un código está formado por s secuencias binarias de longitud n, y es capaz de corregir todos los errores de tamaño inferior o igual a e, se verifica

$$s \leq \frac{2^n}{\sum_{i=0}^{e} \binom{n}{i}} \qquad (3.3.1)$$

En efecto:

Para cada par de palabras, las esferas de radio e deben ser disjuntas. El número de secuencias contenido en cada esfera es

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{e} = \sum_{i=0}^{e} \binom{n}{i}$$

incluída la propia palabra. Luego el número de secuencias incluidas es alguna de las esferas es $s\sum_{i=0}^{e} \binom{n}{i}$ y este número es como máximo igual al número de secuencias posibles 2^n, es decir

$$s\sum_{i=0}^{e} \binom{n}{i} \leq 2^n$$

de donde se obtiene inmediatamente la cota propuesta.

Obsérvese que esta cota constituye una condición necesaria, pero no suficiente para la formación de un código. Por ejemplo, si e=1, n=4, s puede valer 3 según la cota, y sin embargo no hay más que dos palabras cuya distancia sea superior o igual a 3.

Si se admiten errores de tamaño superior a e, no serán corregidos adecuadamente puesto que siempre elegimos el error de menor magnitud.

4.- CODIGOS LINEALES.- CONTROL DE PARIDAD

4.1.- Definiciones

Un código tal que sus palabras contienen dígitos de información y dígitos de control, debe tener éstos elegidos como función de los primeros (si fueran arbitrarios no habría posibilidad de control). Con más precisión, si

son c_1, c_2, ..., c_n los dígitos de una palabra cualquiera, deben cumplirse ciertas relaciones de tipo ecuacional:

$$\begin{aligned} a_{11}c_1 + a_{12}c_2 + \ldots + a_{1n}c_n &= 0 \\ a_{21}c_1 + a_{22}c_2 + \ldots + a_{2n}c_n &= 0 \\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots & \\ a_{r1}c_1 + a_{r2}c_2 + \ldots + a_{rn}c_n &= 0 \end{aligned} \qquad (4.1.1)$$

ó bien, en notación matricial

$$HC' = 0 \qquad (4.1.2)$$

siendo

$$H = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots \\ a_{r1} & \ldots & a_{rn} \end{pmatrix} \; ; \; C = (c_1, \ldots, c_n) \; ; \; 0 = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

La matriz H será denominada matriz de control de paridad y sus elementos son ceros o unos, puesto que las relaciones están construidas sobre el cuerpo de clases residuales módulo 2.

Las soluciones del sistema anterior constituyen un código de control de paridad. Nótese que si el rango de la matriz H es m, hay n-m=k dígitos que pueden elegirse arbitrariamente; son los dígitos de información. Los m dígitos restantes vienen determinados por estos k, y se denominan dígitos de control.

Por ejemplo, sea:

$$H = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

El código de control de paridad resultante está formado por 8 palabras (2^3); hay tres dígitos de información y tres de control en cada una de ellas; se eligen arbitrariamente los tres primeros dígitos (información) y

se determinan los tres restantes (correspondientes a columnas linealmente independientes). Estas ocho palabras son:

000000, 001011, 010101, 011110, 100110, 101101, 110011, 111000.

Observemos que la palabra formada por dígitos cero debe aparecer en todos los códigos de control, de paridad, pues siempre satisface el sistema de ecuaciones.

Supongamos ahora que, dado el código de control de paridad, emitimos una cierta palabra; a su paso por el canal, se suma a dicha palabra una cierta perturbación $E = (e_1, \ldots, e_n)$, siendo

$e_i = 0$, si el canal no cambia el dígito i-ésimo.

$e_i = 1$, si el canal transforma el dígito i-ésimo

Llamaremos a E vector de error.

La secuencia recibida será la suma (siempre módulo 2) de la palabra emitida y la perturbación; es decir, si es $R = (r_1, \ldots, r_n)$ dicha secuencia, se tiene:

$$R = C + E \qquad (4.1.3)$$

A la recepción de la secuencia R puede calcularse el vector llamado corrector (o síndrome) por la fórmula

$$S' = HR' \qquad (4.1.4)$$

Y se cumple:

$$S' = HR' = H(C' + E') = HC' + HE' = HE' \qquad (4.1.5)$$

puesto que C es una palabra y debe ser $HC' = 0$; luego el vector de error determina el corrector, aunque no recíprocamente.

El problema que se presenta ahora es el de determinar, mediante la secuencia recibida, cuál fue la palabra que se emitió. Y se tiene:

$$C = R - E = R + E \qquad (4.1.6)$$

ya que en módulo dos la diferencia coincide con la suma. Bastará, por tanto, determinar el vector de error para lograr conocer la palabra transmitida. Ahora bien, esto no es siempre posible, pues el corector no determina el vector de error de forma unívoca.

4.2.- Teorema

Si es R la secuencia recibida, el conjunto de posibles vectores de error es el conjunto de secuencias que tienen el mismo corrector que R.

En efecto:

Si E es uno de los posibles vectores de error, y fuera

$$HR' \neq HE'$$

se tendría

$$H(R' + E') \neq 0$$

con lo cual R + E no podría ser una palabra. Luego ha de ser

$$HR' = HE' \qquad (4.2.1)$$

c.q.d.

Por ejemplo, si una de las palabras del código es la secuencia recibida, su corrector asociado es el vector nulo. El error adicionado por el canal debe tener corrector nulo y debe ser, por tanto, otra de las palabras del código (la palabra formada por ceros es la que tiene una mayor probabilidad).

Puesto que la relación "tener el mismo corrector" es una relación de equivalencia sobre el conjunto de todas las secuencias, se origina una clasificación en dicho conjunto. Una de las clases (la clase cero) es la formada por las palabras del código. Así resulta que los posibles vectores de error son aquellos que están en la misma clase que la secuencia recibida.

Ninguno de los errores de la misma clase puede ser rechazado con segu-

ridad. Pero, puesto que los errores de canal no son demasiado frecuentes, sí se puede seleccionar entre ellos el error más verosímil, que será aquel con menor número de unos. Es el error que será denominado líder de clase.

Se llama peso de una secuencia al número de unos que aparecen entre sus n dígitos. Así resulta que el líder de clase es la secuencia con menor peso.

Los códigos de control de paridad, es decir, aquellos cuyas palabras son las soluciones de un sistema del tipo (4.1.1), son también llamados códigos lineales o códigos grupo.

La traducción en estos códigos se hará, pues, del siguiente modo: calcular el corrector asociado con la secuencia recibida; determinar el líder de la clase que tiene dicho corrector, y sumar este líder a la secuencia. Así se obtendrá la palabra que con mayor probabilidad ha sido transmitida.

Los problemas que ahora surgen son, fundamentalmente, dos:

a) ¿Cómo debe escogerse la matriz de control de paridad?

b) ¿Cómo puede determinarse el líder de clase, dado el corrector? Para tamaños de palabra pequeños, se puede resolver el problema mediante una búsqueda exhaustiva de todos los líderes, antes de iniciar la traducción; pero cuando esto no es posible aún no hay métodos satisfactorios de búsqueda.

5.- ERRORES SIMPLES.- CODIGOS DE HAMMING.-

5.1.- Construcción.-

Sea E el vector de error adicionado por el canal. Si E tiene unos en las posiciones $j_1, j_2, \ldots, j_e$ y ceros en las n-e posiciones restantes, entonces el corrector $S'=HE'=HU'$ es la suma de las columnas $j_1, \ldots, j_e$ de la matriz de control de paridad.

De este modo, si una columna de la matriz de control de paridad está formada por ceros, un error en dicha posición no será detectado en el corrector.

Y si dos columnas son idénticas, un error simple en cualquiera de las dos posiciones dará lugar al mismo corrector; sólo uno de los dos errores podrá ser corregido. Por tanto, un código lineal no puede corregir todos los errores simples más que si todas las columnas de su matriz de paridad son distintas y no nulas.

Recíprocamente, si la matriz cumple con estas condiciones, cualquier error simple puede ser corregido, pues errores en diferentes posiciones dan lugar a correctores distintos. Podemos, pues, enunciar (si admitimos únicamente la posibilidad de un solo error):

Un código lineal binario es capaz de corregir todos los errores simples si y sólo si todas las columnas de su matriz de paridad son distintas y no nulas.

Para la traducción se empleará el siguiente método: calcular el corrector de la secuencia recibida. Si éste es nulo, se supone que no ha habido error; si es igual a una de las columnas de la matriz de paridad, se corrige el dígito correspondiente a la posición que ocupa la columna. Si, por último, el corrector es no nulo y distinto de todas las columnas de la matriz, no puede efectuarse la traducción (Esta situación sólo es posible cuando ha habido más de un error).

El cálculo de la máxima longitud posible de las palabras, para un código que corrige errores simples, con r dígitos de control, es inmediato, puesto que se reduce a buscar el número máximo de columnas no nulas y distintas que se pueden construir en una matriz de r filas; este número es 2^r-1. La tasa de información viene dada entonces por

$$R = \frac{k}{n} = \frac{2^r-1-r}{2^r-1} = 1 - \frac{r}{2^r-1} \qquad (5.1.1)$$

Haciendo r suficientemente grande puede conseguirse, pues, aproximar la tasa a 1 tanto como se quiera.

5.2.- Número control.-

Cuando la longitud de las palabras es grande (la matriz de paridad tiene gran número de columnas), la comparación del corrector con cada una de las columnas puede necesitar mucho tiempo. Por ello es interesante hacer una ordenación de las columnas tal que, el resultado del corrector indique cuál es la posición en que se ha producido el error sin necesidad de efectuar las comparaciones.

Esto puede conseguirse colocando las columnas de modo que cada una de ellas sea igual al número de orden que ocupa, expresado en base binaria. De este modo el corrector será, en base dos, igual a la posición en que se ha producido el error.

Este tipo especial de corrector será denominado número control.

Por ejemplo:

Para r=3, k=3, n=6, la siguiente matriz de paridad

$$H = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

proporciona un código dado por las palabras:

000000, 001011, 010101, 011110, 100110, 101101, 110011, 111000

Si, en la emisión de la tercera palabra, 010101, se produce un error en el cuarto dígito, se recibirá la secuencia 010001. El cálculo de su corrector da un resultado S = (1 0 0) que es el número 4 expresado en base binaria. La traducción se hará automáticamente cambiando el dígito que se encuentra en la cuarta posición.

BIBLIOGRAFIA:

ASH, R.B.- "Information Theory".- J. Wiley.
BERLEKAMP, E.R.- "Agebraic Coding Theory".- McGraw-Hill.
VAN LINT, J.H.- "Coding Theory".- Springer.