Introducción a la Teoría de Códigos

M.A. García, L. Martínez, T. Ramírez

Facultad de Ciencia y Tecnología. UPV/EHU

Resumen Teórico
Apartado 4 del Tema 3:
Codificación y decodificación para códigos lineales

Mayo de 2017

4 Codificación y decodificación de un código lineal

Como ya hemos indicado, si $G \in \mathrm{Mat}_{k\times n}(\mathbb{F}_q)$ es una matriz generadora del (n,k)-código lineal C, podemos obtener de forma sencilla las palabras de C. Basta considerar $(y_1 \dots y_k) \in \mathbb{F}_q^k$ y calcular $(y_1 \dots y_k)G$. Precisamente, en esta idea se basa el proceso de codificación de palabras. Así, si para construir nuestros mensajes (antes de codificar) vamos a utilizar un diccionario que consta de $\mathbb{F}_q^k$ palabras, podemos identificar cada una de ellas con un elemento $(y_1 \dots y_k) \in \mathbb{F}_q^k$. A continuación, elegimos un (n,k)-código lineal C que en cuanto a su capacidad correctora sea adecuado al canal que vamos a emplear y determinamos una matriz generadora de C, que denotamos por G. Entonces, la codificación de la palabra $(y_1 \dots y_k)$ usando el código elegido será $(y_1 \dots y_k)G$ y esto será precisamente lo que se envíe a través del canal, cuando en el mensaje inicial aparezca la palabra $(y_1 \dots y_k)$.

Ilustramos el proceso de codificación con el siguiente ejemplo:

Ejemplo Supongamos que queremos transmitir una fotografía desde Marte a la Tierra. Hemos mandado un equipo que detecta 8 colores básicos que son

$$L = \{\text{blanco, negro, rojo, amarillo, azul, verde, marrón, violeta}\}.$$

El equipo ha sacado la fotografía y ha dividido la misma en cuadrados muy pequeños y lo que quiere transmitir es el color que aparece en cada uno de ellos. Como las señales que se envían desde Marte solo constan de 0 y 1, se ha identificado cada uno de los colores con una terna de $\mathbb{F}_2^3$, según la siguiente biyección:

$$\begin{array}{rlcl} f: & L & \to & \mathbb{F}_2^3 \\ & \text{blanco} & \mapsto & (1,1,1) \\ & \text{negro} & \mapsto & (0,0,0) \\ & \text{rojo} & \mapsto & (1,0,0) \\ & \text{amarillo} & \mapsto & (0,1,0) \\ & \text{azul} & \mapsto & (0,0,1) \\ & \text{verde} & \mapsto & (1,1,0) \\ & \text{marrón} & \mapsto & (1,0,1) \\ & \text{violeta} & \mapsto & (0,1,1) \end{array}$$

Si mandaramos la información desde Marte utilizando únicamente ternas de $\mathbb{F}_2^3$, no seríamos capaz de detectar si se han producido errores en la transmisión. Esto podría ser un problema porque al estar Marte tan alejado de la Tierra, las señales llegan debilitadas y puede ser que interpretemos de forma errónea el valor que recibimos. Por ello, si únicamente transmitimos desde Marte las ternas asociadas a cada color no tendríamos forma de garantizar que el color que asociamos a un cuadrado determinado sea el verdadero. Pero este problema lo podemos solucionar de forma sencilla usando un código lineal binario de mayor longitud (o sea, que las tuplas usadas para cada color sean más largas) y que tenga dimensión 3, para que conste de 8 palabras.

Por ejemplo, podemos usar el (7,3)-código lineal $C =< 0110100, 0011010, 0001101 >$, que es de longitud 7, dimensión 3 y distancia mínima 3. Empleando C, transmitiríamos $(y_1\ y_2\ y_3)G$, donde $G = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}$. Así, el color blanco se transmitiría como la tupla 0100011, el color negro sería 0000000, el rojo 0110100, etc. Con esta codificación seríamos capaces de detectar hasta 2 errores y corregir 1, ya que el código C elegido tiene distancia mínima $d = 3$. Por tanto, si sabemos que lo más probable sea que se produzca a lo sumo un error en la transmisión de cada 7-tupla, la elección realizada sería adecuada. En cambio, si lo más probable es que se produzcan 2 o más errores, habría que elegir otro código.

Ahora, entre códigos lineales de la misma longitud y misma capacidad de corrección si uno de ellos tiene matriz generadora $G = (I_k|B)$ dada en forma estándar, su codificación será sencilla porque dado $\mathbf{y} = (y_1 \dots y_k) \in \mathbb{F}_q^k$, entonces $\mathbf{y}G = (y_1 \dots y_k|\mathbf{y}B)$ y los k primeros elementos son las coordenadas en la base de G de la palabra emitida $\mathbf{y}G$. Es por este motivo por lo que se prefiere, siempre que sea posible, el uso de matrices generadoras dadas en forma estándar.

Una vez recibido el mensaje el receptor debe verificar si la palabra recibida está o no en el código usado. Si lo está, por el principio de máxima verosimilitud, supondrá que el mensaje recibido es el que le quería enviar el emisor. Si la palabra recibida no pertenece al código usado, significa que se ha producido al menos un error en la transmisión, por lo que procederá a decodificarla, si es posible. El proceso de decodificación precisamente lo que hace es hallar la palabra que ha sido emitida por el emisor, siempre que sea posible. Hay dos métodos de decodificación generales que se emplean para los códigos lineales:

1. Método de decodificación basado en los líderes
2. Método de decodificación mediante síndromes

Ambos métodos, utilizados en códigos lineales, son equivalentes. Esto es, si una palabra recibida $\mathbf{z} \in \mathbb{F}_q^n$ admite una decodificación única y le corresponde como palabra emitida $\mathbf{c} \in C$, esta palabra $\mathbf{c}$ se va a obtener independientemente de cuál sea el método de decodificación que empleemos.

Describimos ahora ambos métodos de decodificación:

4.1 Método de decodificación basado en los líderes

Sea $C \in \mathbb{F}_q^n$ un código lineal. Se define la siguiente relación de equivalencia: $\forall x_1 \cdots x_n, y_1 \cdots y_n \in \mathbb{F}_q^n$

$$x_1 \cdots x_n \sim y_1 \cdots y_n \iff x_1 \cdots x_n - y_1 \cdots y_n \in C.$$

Observamos que

$$\begin{array}{rcl}[x_1\cdots x_n] & = & \{y_1\cdots y_n\in\mathbb{F}_q^n | x_1\cdots x_n\sim y_1\cdots y_n\}\\ & = & \{y_1\cdots y_n\in\mathbb{F}_q^n | x_1\cdots x_n - y_1\cdots y_n\in C\}\end{array}.$$

Entonces, si recibimos la palabra $z_1\cdots z_n\in\mathbb{F}_q^n$, calculamos $[z_1\cdots z_n]$ y como el número de errores producidos en la transmisión lo suponemos mínimo por el principio de máxima verosimilitud, la decodificamos como $z_1\cdots z_n - y_1\cdots y_n$, donde $y_1\cdots y_n\in[z_1\cdots z_n]$ y es de peso mínimo. Si hay un único $y_1\cdots y_n\in[z_1\cdots z_n]$ de peso mínimo, llamaremos a $y_1\cdots y_n$ **líder** de $[z_1\cdots z_n]$ y este líder es precisamente el error cometido en la transmisión. Si hay varias $y_1\cdots y_n\in[z_1\cdots z_n]$ con el mismo peso mínimo diremos que $z_1\cdots z_n$ no admite decodificación única.

4.2 Método de decodificación mediante síndromes

Sea $C\in\mathbb{F}_q^n$ un código lineal con matriz de control H. Dada $z_1\cdots z_n\in\mathbb{F}_q^n$ se llama **síndrome de** $z_1\cdots z_n\in\mathbb{F}_q^n$ a $S(z_1\cdots z_n)=(z_1\cdots z_n)H^t$. Observamos que las palabras del código C satisfacen que su síndrome es $\mathbf{0}$. Además, podemos definir la relación de equivalencia: $\forall x_1\cdots x_n, y_1\cdots y_n\in\mathbb{F}_q^n$

$$x_1\cdots x_n\mathfrak{R}y_1\cdots y_n \iff S(x_1\cdots x_n)=S(y_1\cdots y_n)$$

Entonces, si $x_1\cdots x_n\mathfrak{R}y_1\cdots y_n$, se tiene $S(x_1\cdots x_n)=S(y_1\cdots y_n)$ lo que implica $S(x_1\cdots x_n - y_1\cdots y_n)=\mathbf{0}$, esto es, $x_1\cdots x_n - y_1\cdots y_n\in C$. Esto nos proporciona el siguiente método de decodificación:

Si recibimos la palabra $z_1\cdots z_n\in\mathbb{F}_q^n$, determinamos $S(z_1\cdots z_n)$ y decodificamos $z_1\cdots z_n$ como $z_1\cdots z_n - y_1\cdots y_n$ siendo $y_1\cdots y_n\in\mathbb{F}_q^n$ de peso mínimo tal que $S(z_1\cdots z_n)=S(y_1\cdots y_n)$.

Ejemplo Sea C el $(6,3)$-código binario con matriz de control

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

Si se recibe la tupla $\mathbf{v}=110101$, entonces $S(110101)=(0\ 0\ 1)$. Pero $S(100000)=(0\ 0\ 1)$, así que la tupla se corrige a

$$u=\mathbf{v}-\mathbf{e}=010101.$$

En cambio, si recibimos $\mathbf{y}=100001$ tenemos que $S(\mathbf{y})=(1\ 1\ 1)$. Pero no hay ningún vector de peso 1 con este síndrome y sí al menos dos de peso 2: 100001 y 001100 con el mismo síndrome, por lo que $\mathbf{y}$ no tiene decodificación única, pues podríamos decodificarla como 000000 ó 101101.