Introducción a la Teoría de Códigos

M.A. García, L. Martínez, T. Ramírez

Facultad de Ciencia y Tecnología. UPV/EHU

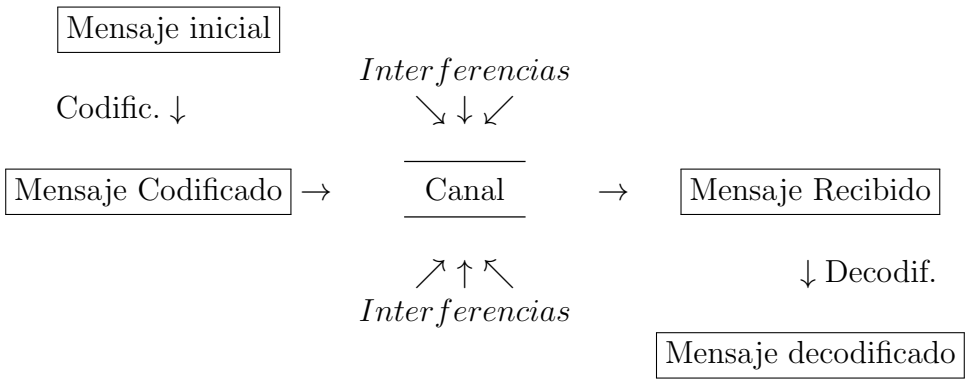
Resumen Teórico
Apartado 2 del Tema 2:
Códigos detectores y correctores de errores

Mayo de 2017

2 Códigos correctores y detectores de errores

Como se ha indicado en la sección anterior, el problema con el que nos enfrentamos es que deseamos enviar un mensaje a un receptor a través de un canal que puede ser poco fiable, esto es, el mensaje que le llega no tiene que ser el emitido y puede contener errores de transmisión. Lo que nos interesa es que el receptor, tras recibir el mensaje, pueda detectar si el mensaje que le ha llegado es el original y si no lo es, intentar recuperar el mensaje inicial.

De forma esquemática, resumimos las fases que componen todo el proceso como sigue:



Buscamos que el mensaje decodificado sea lo más parecido posible al mensaje codificado. Es más, nos interesa haber elegido bien el código de acuerdo al canal utilizado para poder tener una probabilidad muy alta de que el mensaje codificado y el decodificado sean iguales para que pueda el receptor deducir cuál era el mensaje inicial a partir del decodificado, sabiendo el código que se ha empleado.

En el estudio de los códigos correctores y detectores de errores usaremos algunas definiciones básicas que conviene que precisemos a qué se refieren cuando las empleemos. Por ello, a continuación vamos a formalizar matemáticamente conceptos que aparecen de forma frecuente, como los siguientes:

1. **Alfabeto**: Es un conjunto finito de elementos, llamados letras. Así, matemáticamente un alfabeto es un conjunto

$$A = \{a_1, \ldots, a_m\},$$

siendo cada $a_i \in A$ una letra del alfabeto A y $m \in \mathbb{N}$ el cardinal del alfabeto A.

2. **Palabra**: Fijado un alfabeto A, una palabra sobre el alfabeto A es un elemento formado por la concatenación de un número finito de letras de A. Salvo que se indique lo contrario, denotaremos las palabras por letras minúsculas del alfabeto latino en negrita. Por ejemplo, $\mathbf{x} = x_1 \ldots x_n$, donde $x_i \in A$ para

CC BY NC SA

$i = 1, \ldots, n$ es una palabra sobre A que se ha formado al poner de forma consecutiva las letras x_1,...,x_n. En muchos textos a estas palabras se les denomina "*palabras positivas*" para distinguirlas de aquellas en las que las letras x_i pueden venir afectadas del exponente -1.

3. Ω_A: Es el semigrupo libre con base A, esto es, el conjunto de todas las palabras (positivas) que se pueden formar con las letras del alfabeto A.

4. **Código**: Es un conjunto de palabras sobre un mismo alfabeto, es decir, $C \subseteq \Omega_A$. A los elementos de C se les llama palabras del código C.

5. **Longitud de una palabra** $l(\mathbf{x})$: Número de letras que tiene una palabra. Si $\mathbf{x}= x_1 \ldots x_n$, entonces $l(\mathbf{x}) = n$.

6. **Igualdad de palabras**: Dos palabras sobre el mismo alfabeto A $\mathbf{x}$ e $\mathbf{y}$ son iguales si tienen la misma longitud y coinciden los valores de cada letra en cada posición. Esto es,

$$\mathbf{x} = \mathbf{y} \Longleftrightarrow \mathbf{x} = x_1 \cdots x_n, \quad \mathbf{y} = y_1 \cdots y_n \text{ y } x_i = y_i, \forall i = 1, \ldots, n.$$

7. **Código de bloque**: Dado un código $C \subseteq \Omega_A$, se dice que C es un código de bloque si todas las palabras de C tienen la misma longitud n. También se suele llamar código de longitud n.

8. **Distancia de Hamming**: Dadas $\mathbf{x}$ e $\mathbf{y}$ dos palabras de la misma longitud se llama distancia de Hamming entre $\mathbf{x}$ e $\mathbf{y}$ al número de componentes en que difieren ambas palabras. Esto es, si $\mathbf{x}= x_1 \ldots x_n$ e $\mathbf{y}= y_1 \ldots y_n$,

$$d(\mathbf{x}, \mathbf{y}) = |\{i \in \{1, \ldots, n\} \,|\, x_i \neq y_i\}|,$$

donde $d(\mathbf{x}, \mathbf{y})$ denota la distancia de Hamming entre las palabras $\mathbf{x}$ e $\mathbf{y}$).

9. **Distancia mínima de un código de bloque** C **de longitud** n: Dado $C \subseteq \Omega_A$ tal que C es un código de bloque, se llama distancia mínima de C a la menor de las distancias entre palabras diferentes de C, esto es,

$$d(C) = \min\{d(x_1 \ldots x_n, y_1 \ldots y_n) \,|\, \begin{array}{l} x_1 \ldots x_n, y_1 \ldots y_n \in C, \\ x_1 \ldots x_n \neq y_1 \ldots y_n\} \end{array}$$

10. **Peso de una palabra**: Si $\mathbf{x} \in \mathbb{F}_q^n$, se llama peso de $\mathbf{x}$, y se denota por $w(\mathbf{x})$, al número de componentes no nulas de $\mathbf{x}$.

11. **Peso mínimo de un código**: Si $C \subseteq \mathbb{F}_q^n$, se llama peso mínimo de C, y se denota por $w(C)$, a

$$w(C) = \min\{w(\mathbf{x}) \,|\, \mathbf{x} \in C - \{\mathbf{0}\}\}.$$

12. **Detectar hasta** t **errores**: Saber que se ha producido t fallos en la transmisión, esto es, que si trabajamos con un código C que detecta t errores, entonces cualquier palabra $\mathbf{y}$ que se obtenga de una palabra de $\mathbf{c} \in C$ cambiando el valor que figura en $\mathbf{c}$ en a lo sumo t posiciones, resulta que $\mathbf{y} \notin C$.

CC BY NC SA

13. **Corregir hasta t errores**: Ser capaz de recuperar la palabra original cuando se han producido t fallos en la recepción de la misma, es decir que si $\mathbf{y} \notin C$ se ha obtenido a partir de $\mathbf{c} \in C$ cambiando el valor que figura en $\mathbf{c}$ en a lo sumo t posiciones, entonces la única palabra de C que dista de $\mathbf{y}$ a lo sumo t es la propia $\mathbf{c}$. En este caso diremos que $\mathbf{c}$ es la decodificación (única) de $\mathbf{y}$.

14. **Principio de máxima verosimilitud**: En las transmisiones es más probable que se produzca siempre el menor número de fallos posible. Seguiremos este principio en todo el curso.

Ejemplos

1. El cuerpo finito $\mathbb{F}_q$, con $q = p^t$, siendo p número primo y $t \in \mathbb{N}$, es un ejemplo de alfabeto que utilizaremos en este curso. Si $q = 2$, esto es, el alfabeto es $\mathbb{F}_2 = \{0, 1\}$ y tomamos un código C de $\Omega_{\mathbb{F}_2}$, se dice que el código C es binario.

2. $\mathbf{x}$=01010 es una palabra de longitud 5 usando como alfabeto $\mathbb{F}_2$ o en general $\mathbb{F}_q$.

3. Si tomamos $A = \mathbb{F}_2 = \{0, 1\}$ y

$$C = \{\mathbf{x} \in \Omega_A \mid \mathbf{x} \text{ contiene un número par de “1”}\},$$

entonces C es un código binario. Una palabra de C es $\mathbf{x}$= 011001010 y la longitud de $\mathbf{x}$ es 9. C no es un código de bloques porque las palabras $\mathbf{y}$= 011 y $\mathbf{z}$= 1010 están en C y tienen longitud distinta. Por otro lado, la palabra $\mathbf{y}$= 011001000 $\in \Omega_A$ no pertenece a C. Otra forma de caracterizar los elementos de C es la siguiente:

$$C = \{\mathbf{x} = x_1 \dots x_m \in \Omega_A \mid m \in \mathbb{N}, \sum_{i=1}^{m} x_i \equiv 0 \bmod 2\}.$$

4. Los siguientes ejemplos son ejemplos conocidos de códigos de bloque:

   - Código ASCII: Es un código que se emplea para transmitir la información desde el teclado del ordenador a la CPU. A cada letra y símbolo del teclado se le asigna un número entre 0 y 127. Este número se representa en el sistema binario como una 7-tupla, que se completa con un octavo dígito (0, ó 1) de forma que la 8-tupla resultante tenga un número par de “1”. Por ejemplo a la letra A le corresponde el 65 que se representa por la 7-tupla 1000001 y se añade a esta 7-tupla un “0” de forma que la 8-tupla resultante tenga un número par de 1, dando lugar en este caso a la 8-tupla 10000010 que es la que correspondería a la letra A.
   - Código ISBN: Es un código que identifica de forma única a los libros publicados. A cada libro se le asocia un número de diez cifras $a_1 \dots a_{10}$ en la que las nueve primeras, que toman valores entre 0 y 9, dan información sobre el libro (país, editorial, título,...) y la última se obtiene exigiendo

CC BY NC SA

que $\sum_{i=1}^{10} ia_i \equiv 0 \bmod 11$. Si a_{10} toma el valor 10, se sustituye éste por X, por lo que $a_{10} \in \{0, 1, \ldots, 9, X\}$. Por ejemplo, una palabra que forma parte de este código ISBN es 0-19-853803-0.

- Código EAN: Es un código que permite identificar de forma única a los productos que se vendenen Europa. A cada producto se le asocia un número de trece cifras $a_0 \ldots a_{12}$, con $a_i \in \{0, 1, \ldots, 9\}$ para $i = 0, 1, \ldots, 12$, donde los 3 primeros dígitos representan el código del país en donde radica la empresa que lo comercializa, los 4 o 5 siguientes son el código de empresa, que identifica al propietario de la marca y es asignado por la asociación de fabricantes y distribuidores (AECOC). Los 5 o 4 siguientes son el código de producto y el último dígito (a_{12}) es un dígito, llamado de control, que se obtiene de forma que $3\sum_{i=0}^{5} a_{2i+1} + \sum_{i=0}^{6} a_{2i} \equiv 0 \bmod 10$.

CC BY NC SA