

Teoría de la Información

Enrique Alexandre

Dpto. de Teoría de la Señal y Comunicaciones

Curso 2021/2022

Índice

1. Introducción
 - Medida de la información
 - Características de los códigos fuente
2. Codificación Huffman
 - Introducción
 - Ejemplo
3. Codificación LZW
 - Definición
 - Pseudocódigo
 - Ejemplo
 - Decodificación
 - Ejemplo

Introducción

Medida de la información

- Se define la cantidad de información como:

$$I = -\log_2(p_k)$$

siendo p_k la probabilidad de aparición de cada uno de los posibles símbolos de la fuente.

- Se define la **entropía** de una fuente como:

$$H = - \sum_{k=0}^{N-1} p_k \cdot \log_2(p_k)$$

La entropía se mide en bits, e indica cuántos bits por símbolo como mínimo son necesarios para codificar la fuente.

Introducción

Codificación de fuente

- **Objetivo:** Reducir el número de bits necesarios para representar la salida de una fuente.
- **Teorema de Shannon:** El número mínimo de bits por símbolo nunca va a ser inferior al valor de la entropía de la fuente.
- Códigos fuente:
 - Codificación Huffman: creada por David Huffman.
 - Codificación LZW: creada por Abraham Lempel, Jacob Ziv y Terry Welch.

Introducción

Características de los códigos fuente

- **Longitud media:** Se define como la longitud promedio de las palabras del código

$$\bar{L} = \sum_{i=1}^M p_i n_i$$

- **Tasa de compresión:** Se define como la relación frente a un código de longitud fija:

$$\Gamma = \frac{\lceil \log_2 M \rceil}{\bar{L}}$$

- **Eficiencia:** Mide lo cerca que estamos del límite teórico de la entropía:

$$\eta = \frac{H(x)}{\bar{L}} \leq 1$$

Codificación Huffman

Introducción

- Símbolos menos probables -> Se codifican con más bits
- Símbolos más probables -> Se codifican con menos bits
- El código se construye de forma iterativa en forma de árbol

Codificación Huffman

Ejemplo

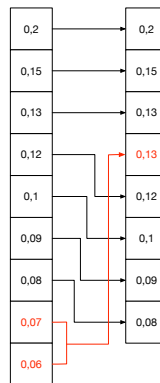
- Supongamos una fuente con nueve símbolos con las siguientes probabilidades:

$$\mathbf{p} = (0,2, 0,15, 0,13, 0,12, 0,1, 0,09, 0,08, 0,07, 0,06)$$

Codificación Huffman

Ejemplo

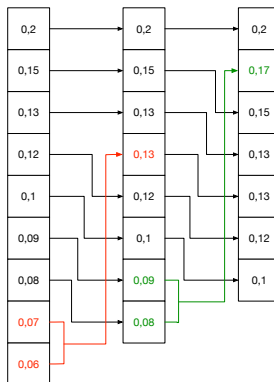
- Agrupo los dos símbolos menos probables en uno cuya probabilidad será la suma y reordeno:



Codificación Huffman

Ejemplo

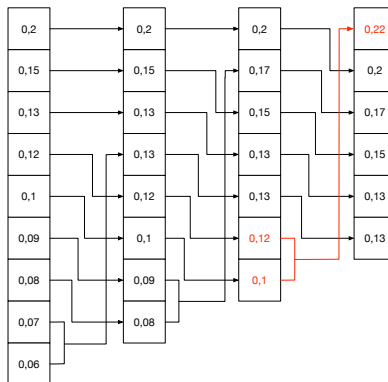
- Repito el proceso hasta que sólo queden dos símbolos:



Codificación Huffman

Ejemplo

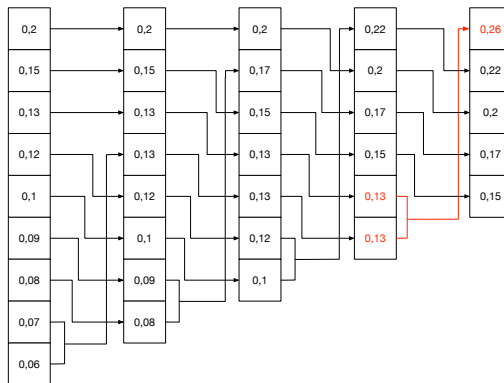
- Repito el proceso hasta que sólo queden dos símbolos:



Codificación Huffman

Ejemplo

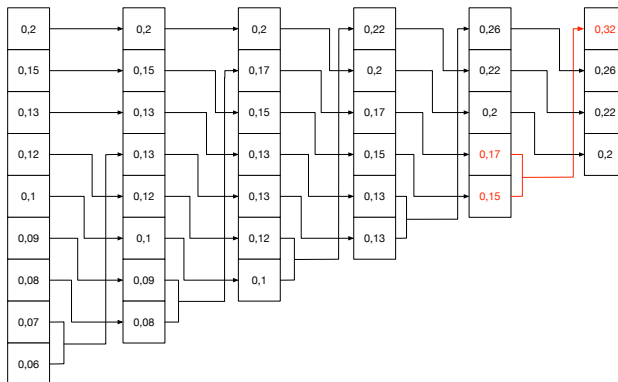
- Repito el proceso hasta que sólo queden dos símbolos:



Codificación Huffman

Ejemplo

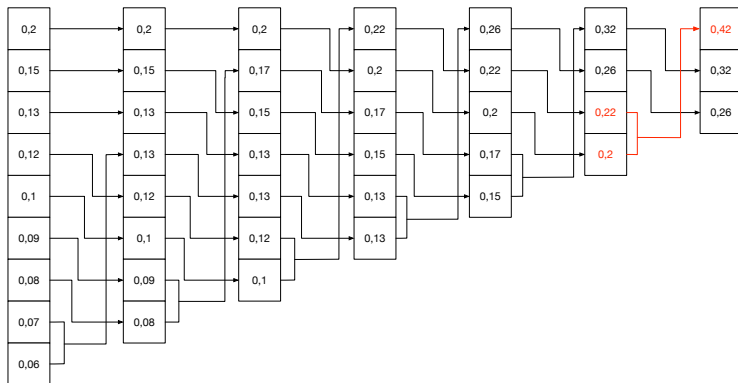
- Repito el proceso hasta que sólo queden dos símbolos:



Codificación Huffman

Ejemplo

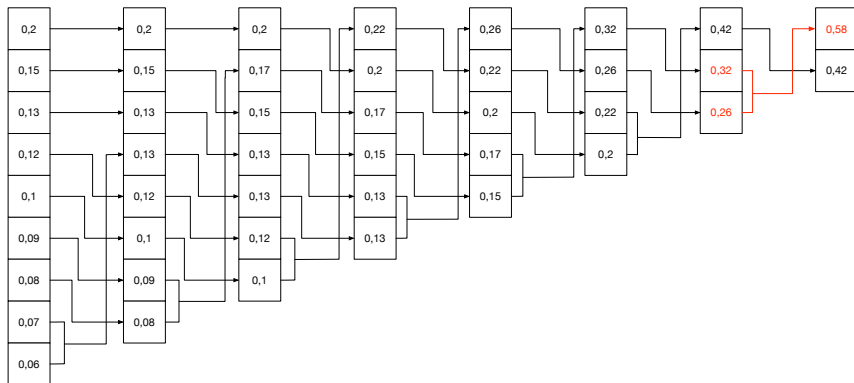
- Repito el proceso hasta que sólo queden dos símbolos:



Codificación Huffman

Ejemplo

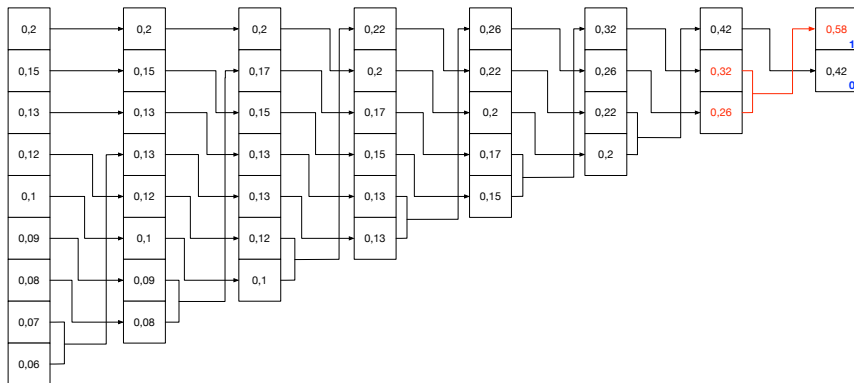
- Repito el proceso hasta que sólo queden dos símbolos:



Codificación Huffman

Ejemplo

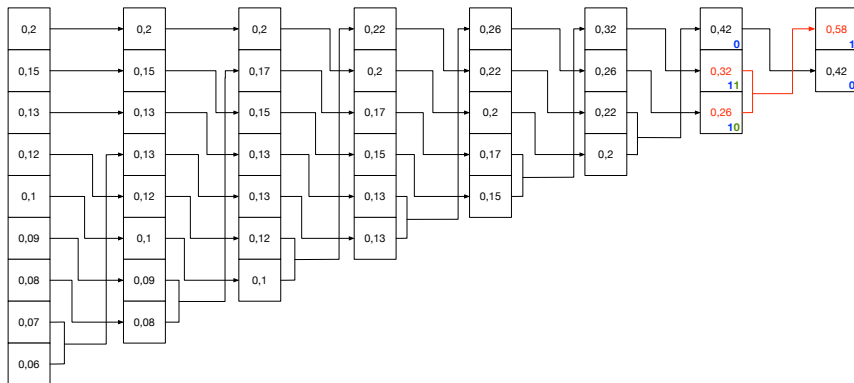
- Asigno 'o' a un símbolo y 'r' al otro:



Codificación Huffman

Ejemplo

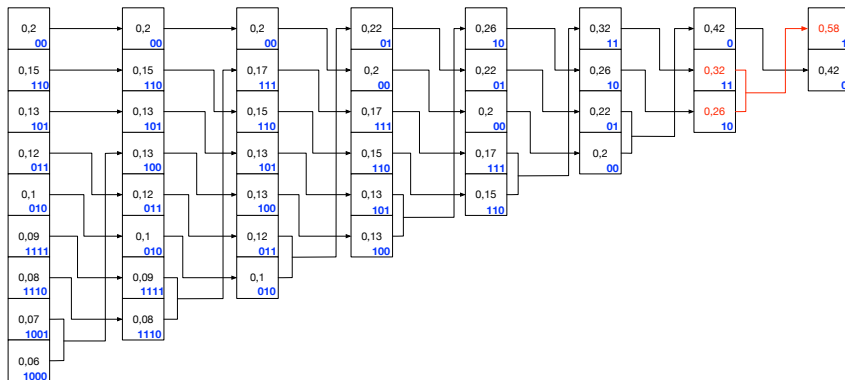
- Voy hacia atrás, añadiendo 'o' ó 'r' según corresponda:



Codificación Huffman

Ejemplo

- Completo el proceso:



Codificación Huffman

Ejemplo

- Así, finalmente:

0, 2	00
0, 15	00
0, 13	00
0, 12	00
0, 1	00
0, 09	00
0, 2	00
0, 2	00
0, 2	00

Codificación Huffman

Ejemplo

- Longitud media:

$$\bar{L} = 2 \cdot 0,2 + 3 \cdot (0,15 + 0,13 + 0,12 + 0,1) + 4 \cdot (0,09 + 0,08 + 0,07 + 0,06) = 3,1$$

- Se puede demostrar que se cumple que:

$$H \leq \bar{L} \leq H + 1$$

Codificación LZW

Definición

- Tienen la ventaja de que no necesitamos conocer las probabilidades a priori de los símbolos.
- Se utiliza, por ejemplo, en los formatos de imagen GIF y TIFF.

Codificación LZW

Pseudocódigo

```

CADENA = cadena vacía
WHILE quedan caracteres por codificar DO
    CHARACTER = coger el siguiente carácter
    IF CADENA+CHARACTER está en el diccionario
        CADENA = CADENA+CHARACTER
    ELSE
        código correspondiente a CADENA -> SALIDA
        Añadir CADENA+CHARACTER al diccionario
        CADENA = CHARACTER
    END
END
código para CADENA -> SALIDA

```

Codificación LZW

Ejemplo

- Supongamos que partimos del siguiente diccionario:

Entrada	Código
a	0
b	1
n	2

- El mensaje de entrada al codificador es: “banana”

Codificación LZW

Ejemplo

- Siguiendo el proceso del pseudocódigo, vamos construyendo el diccionario y codificando el mensaje:

CADENA	CARACTER	¿En el Diccionario?	Al Diccionario	Salida
	b	Si		
b	a	No	3 - ba	1
a	n	No	4 - an	0
n	a	No	5 - na	2
a	n	Sí		
an	a	No	6 - ana	4
a				0

- Al final, el mensaje codificado es: 1 0 2 4 0.

Codificación LZW

Decodificación

```
CODIGO_1 = Leer primer código del mensaje
Traducción de CODIGO_1 -> SALIDA
WHILE queden caracteres por decodificar
    CODIGO_2 = Leer siguiente código
    CADENA = traducción de CODIGO_2
    CADENA -> SALIDA
    CARACTER = Primer carácter de CADENA
    Añadir (Traducción de CODIGO_1)+(CARACTER) al diccionario
    CODIGO_1 = CODIGO_2
END
```


Codificación LZW

Ejemplo

- Siguiendo el proceso del pseudocódigo, vamos reconstruyendo el diccionario y decodificando el mensaje:

CODIGO ₁	CODIGO ₂	CADENA	CARACTER	Salida	Dicc.
1				b	
1	0	a	a	a	3 - ba
0	2	n	n	n	4 - an
2	4	an	a	an	5 - na
4	0	a	a	a	6 - ana