Comunicaciones Digitales

Práctico 6

Teoría de la información: teorema de codificación de Shannon

Cada ejercicio comienza con un símbolo el cuál indica su dificultad de acuerdo a la siguiente escala: ✦ básica, ★ media, ✷ avanzada, y ✹ difícil.

✦ Ejercicio 1

Un teclado numérico tiene los números $0, 1, 2 \ldots 9$. Se asume que cada tecla es utilizada en forma equiprobable. Calcule con qué cadencia deben ser oprimidas las mismas para generar un flujo de información de 2 bit/s.

✦ Ejercicio 2

Calcular la tasa de información de una fuente telegráfica que utiliza dos símbolos, el punto y la raya. El punto tiene una duración de 0.2s. La raya tiene el doble de duración pero es la mitad de probable.

✦ Ejercicio 3

Para una fuente binaria:

(a) Mostrar que la entropía, H, es máxima cuando la probabilidad de enviar un 1 es igual a la probabilidad de enviar un 0.

(b) Hallar el valor máximo de la entropía.

★ Ejercicio 4

Se tiene una fuente S con símbolos $\{s_1, s_2, \ldots, s_6\}$, con probabilidades 0.4, 0.3, 0.1, 0.1, 0.06 y 0.04 respectivamente.

(a) Hallar la entropía de la fuente.

(b) Considerar una codificación binaria con largo de palabra fijo. Indicar el mínimo largo de palabra para que la codificación sea adecuada.

(c) Considerar una codificación de Huffman. Calcular el largo medio del código obtenido.

✦ Ejercicio 5

Considerar un canal con la propiedad de que x_i y y_j son estadísticamente independientes para todo i, j. Mostrar que

$$H(X|Y) = H(X) \text{ y } I(X, Y) = 0$$

⋆ Ejercicio 6

Se quiere codificar un archivo que constará únicamente de números del 0 al 9, sin saltos de línea, espacios o caracteres de puntuación.

(a) ¿Cuántos bits se requieren para codificar cada dígito? Proponga un código de largo fijo.

Si basado en un análisis de múltiples archivos de similares características, ahora se sabe que la probabilidad de ocurrencia de cada uno de los caracteres es la siguiente:

Caracter	Probabilidad
0	$0,20$
1	$0,10$
2	$0,07$
3	$0,07$
4	$0,06$
5	$0,10$
6	$0,06$
7	$0,07$
8	$0,07$
9	$0,20$

(b) Calcule la entropía de la fuente.

(c) Pronga un código que optimice el uso de espacio en disco a la hora de almacenar este tipo de archivos. ¿Cómo es su largo medio respecto a la entropía de la fuente?

(d) ¿Cómo podemos acercarnos más a la cota de Shannon? ¿Cómo debe ser el archivo para poder hacerlo?

Si las probabilidades de ocurrencia de los dígitos fueran ahora:

Caracter	Probabilidad
0	$0,2500$
1	$0,0625$
2	$0,0625$
3	$0,0625$
4	$0,0625$
5	$0,0625$
6	$0,0625$
7	$0,0625$
8	$0,0625$
9	0.2500

(e) ¿Cambia la entropía de la fuente?

(f) Vualva a proponer un código que optimice el uso de espacio en disco.

(g) ¿Cómo es el largo medio del nuevo código respecto a la entropía de la fuente? ¿Por qué?

Solución

Ejercicio 1

La tasa de símbolos r se relaciona con la entropía H y con la tasa de información R según la expresión:

$$R = rH$$

La probabilidad de cada una de las teclas es : $P(Tecla) = \frac{1}{10}$. Por lo que la entropía de la fuente S={0,1,2, ... 9} es:

$$H(S) = \sum_{x_i S} p(x_i) log_2 \frac{1}{p(x_i)} = \sum_{x_i S} \frac{1}{10} log_2 10 = log_2 10 \; bits/simb$$

Despejando r se tiene:

$$r = \frac{R}{H} = \frac{2 \; bits/seg}{log_2 10 \; bits/simb} = 0,6 \, simb/seg$$

Ejercicio 2

La cadencia media de símbolos esta dada por:

$$r = \frac{1}{0,2 \frac{seg}{simb} . \frac{2}{3} + 0,4 \frac{seg}{simb} . \frac{1}{3}} = \frac{30}{8} \; \frac{simb}{seg}$$

La entropía es:

$$H = \frac{2}{3} log_2 \frac{3}{2} + \frac{1}{3} log_2 3 = 0,9183 \; \frac{bits}{simb}$$

Por tanto,

$$R = r.H = 3,44 \; \frac{bits}{seg}$$

Ejercicio 3

(a) Sea p la probabilidad de enviar un uno. La probabilidad de enviar un 0 será entonces $1 - p$, y la entropía queda planteada de la siguiente forma:

$$H(p) = p.log \frac{1}{p} + (1-p).log \frac{1}{1-p}$$

Para saber cuándo es máxima sólo falta derivar respecto a p:

$$H' = log \frac{1}{p} - \frac{p^2}{p^2} - log \frac{1}{1-p} + (1-p). \frac{1-p}{(1-p)^2} = log \frac{1-p}{p}$$

En el máximo se debe cumplir $H' = 0$

$$log \frac{1-p}{p} = 0$$

Tenemos que $\frac{1-p}{p} = 1$ y entonces:

$$p = \frac{1}{2}$$

Falta verificar que efectivamente es un máximo ($H'' < 0$):

$$H''(p) = \frac{p}{1-p} . \frac{-p-(1-p)}{p^2} = \frac{-1}{(1-p)p} = -4$$

(b)

$$H(\frac{1}{2}) = 2.\frac{1}{2}.log2 = log_2 2\frac{bits}{simb} = 1\frac{bit}{simb}$$

Ejercicio 4

(a) H(S) $= P_1 \log_2 \frac{1}{s_1} + P_2 \log_2 \frac{1}{s_2} + P_3 \log_2 \frac{1}{s_3} + P_4 \log_2 \frac{1}{s_4} + P_5 \log_2 \frac{1}{s_5} + P_6 \log_2 \frac{1}{s_6} = 2.1435$ bits/símbolo.

(b) El mínimo largo de palabra de código para codificar 6 símbolos es 3 bits.

(c) Considerando la extensión que se muestra en el cuadro 1, el largo de código resulta: $L = 1 \times 0.4 + 2 \times 0.3 + 3 \times 0.1 + 4 \times 0.1 + 5 \times 0.06 + 5 \times 0.04 = 2.2$ bits/símbolo. Nótese que con este código se llega muy cerca de la entropía de la fuente. Además, aunque la extensión no es única, el largo no depende de la extensión elegida.

Fuente original			Fuente reducida							
Símbolos	Prob.	Código	S_1		S_2		S_3		S_4	
s_1	0.4	1	0.4	1	0.4	1	0.4	1	$\circ$0.6	0
s_2	0.3	00	0.3	0	0.3	00	0.3	00$\circ$	0.4	1
s_3	0.1	011	0.1	011	$\star$0.2	010$\diamond$	$\diamond$0.3	01$\circ$		
s_4	0.1	0100	0.1	0100$\star$	0.1	011$\diamond$				
s_5	0.06	01010*	*0.1	0101$\star$						
s_6	0.04	01011*								

Cuadro 1: Síntesis de un código de Huffman.

Ejercicio 5

Aplicando la definición:

$$H(X|Y) = \sum_{X,Y} P(x_i, y_j).log_2 \frac{1}{P(x_i|y_j)}$$

Como x_i y y_j son independientesse cumple que$P(x_i, y_j) = P(x_i).P(y_j)$ y que $P(x_i|y_j) = P(x_i)$, por lo que:

$$H(X|Y) = \sum_{Y}\sum_{X} P(x_i)P(y_j).log_2 \frac{1}{P(x_i)} = \sum_{Y} P(y_j) \sum_{X} P(x_i).log_2 \frac{1}{P(x_i)}$$

$$H(X|Y) = \sum_{X} P(x_i).log_2 \frac{1}{P(x_i)} = H(X)$$

Para la segunda parte, aplicando la definición:

$$I(X,Y) = \sum_{X,Y} P(x_i, y_j).log_2 \frac{P(x_i|y_j)}{P(x_i)}$$

$$I(X,Y) = \sum_{X,Y} P(x_i)P(y_j).log_2 \frac{P(x_i)}{P(x_i)} = \sum_{X,Y} P(x_i)P(y_j).0 = \sum_{X,Y} 0 = 0$$

Ejercicio 6

(a) Se requieren al menos 4 bits, que serán capaces de codificar hasta $2^4 = 16$ sínbolos distintos. Un codigo posible sería:

Caracter	Código
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001

(b) La entropía de la fuente vale

$$H\{S\} = \sum_i p_i log_2 \frac{1}{p_i} \approx 3.154 \text{ bits/símbolo.}$$

(c) A partir del algoritmo de Huffman es posible lograr el siguiente código:

Símbolos	Prob.	Código
0	$0,20$	10
1	$0,10$	0000
2	$0,07$	0100
3	$0,07$	0101
4	$0,06$	0110
5	$0,10$	0001
6	$0,06$	0111
7	$0,07$	0010
8	$0,07$	0011
9	$0,20$	11

Y es posible ver que su largo medio vale:

$$\bar{L}\{C\} = 3.2 \text{ bits.}$$

(d) Debemos tomar bloques de n dígitos, con n grande, y codificar como si cada uno de esos bloques fuera un nuevo símbolo. A esto lo llamamos una extensión de la fuente.
La entropía de esta nueva fuente será $nH\{S\}$.

(e) Sí, cambia. La entropía de la nueva fuente es

$$H\{S\} = \sum_i p_i log_2 \frac{1}{p_i} = 3.0 \text{ bits/símbolo.}$$

(f) Volvemos a implementar el algoritmo de Huffman para lograr el siguiente código:

Símbolos	Prob.	Código
0	$0,2500$	00
1	$0,0625$	1000
2	$0,0625$	1001
3	$0,0625$	1010
4	$0,0625$	1011
5	$0,0625$	1100
6	$0,0625$	1101
7	$0,0625$	1110
8	$0,0625$	1111
9	0.2500	01

Y es posible ver que el nuevo largo medio vale:

$$\bar{L}\{C\} = 3 \text{ bits.}$$

(g) En este caso el largo medio del código coincide con la entropía de la fuente. Sucede así por cómo fueron asignadas las probabilidades de ocurrencia de los símbolos: la información propia (*self-information*) de cada uno coincide con un número entero de bits.