Parcial 1 (12 Abril 2018), preguntas y respuestas

Informació i Seguretat (Universitat Autònoma de Barcelona)

INFORMACIÓ I SEGURETAT
12 d'abril de 2018

Nom i cognoms (en MAJÚSCULES): ______________________ NIU: ________ Grup: ___

****** Cal que **justifiqueu convenientment** totes les respostes.

****** $\log 3 = 1.58$, $\log 5 = 2.32$, $\log 7 = 2.80$, $\log 23 = 4.52$.

1. (2.5 punts, 1+0.75+0.75) Una font emet a l'atzar, i de manera equiprobable, símbols quaternaris. Volem enviar els símbols per un canal binari i fem servir el següent codi $S = \{00, 01, 10, 11\}$. Cada cop que enviem un 0 a l'entrada del canal obtenim un 0, però la meitat de vegades que enviem un 1 a l'entrada, obtenim un 1 a la sortida, i la meitat de vegades, un 0. Per exemple, si enviem l'entrada 10, a la sortida del canal podem obtenir 10 o 00.

   (a) Quina informació mitjana ens donen els elements de la sortida del canal, $R = \{00, 01, 10, 11\}$?

   (b) Quina és la informació de l'element enviat si sabem que hem rebut 11 a la sortida del canal?

   (c) Quina informació tenim dels valors enviats segons els obtinguts a la sortida del canal?

   Solució:

$p(R_j\|S_i)$	00	01	10	11
00	1	0	0	0
01	$\frac{1}{2}$	$\frac{1}{2}$	0	0
10	$\frac{1}{2}$	0	$\frac{1}{2}$	0
11	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

$p(R_j, S_i)$	00	01	10	11
00	$\frac{1}{4}$	0	0	0
01	$\frac{1}{8}$	$\frac{1}{8}$	0	0
10	$\frac{1}{8}$	0	$\frac{1}{8}$	0
11	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
$p(R_j)$	$\frac{9}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$

$p(S_i\|R_j)$	00	01	10	11
00	$\frac{4}{9}$	0	0	0
01	$\frac{2}{9}$	$\frac{2}{3}$	0	0
10	$\frac{2}{9}$	0	$\frac{2}{3}$	0
11	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{1}{3}$	1

   (a) La distribució final de probabilitats és $\{\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}\}$. Per tant, la informació mitjana que ens donen els elements de la sortida del canal és $H(R) = H(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}) = \log(16) - \frac{24}{16}\log(3) = 4 - \frac{3}{2}1.58 = 4 - 2.37 = 1.63$.

   (b) Si a la sortida del canal obtenim 11, aleshores l'element a l'entrada només pot ser 11. Així, si a la sortida del canal obtenim 11, la incertesa de l'element de l'entrada és zero i per tant la informació de l'element de l'entrada és zero. Si ho mirem a les taules, tenim que $H(S|11) = H(0, 0, 0, 1) = 0$.

   (c) Apliquem la fórmula $I(S, R) = H(R) - H(R|S)$. Tenim que $H(R) = 1.63$ i $H(R|S) = \frac{1}{4}H(1, 0, 0, 0) + \frac{1}{4}H(\frac{1}{2}, \frac{1}{2}, 0, 0) + \frac{1}{4}H(\frac{1}{2}, 0, \frac{1}{2}, 0) + \frac{1}{4}H(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}) = 1$. Per tant, $I(S, R) = 1.63 - 1 = 0.63$.
   També es pot obtenir el resultat aplicant la fórmula $I(S, R) = H(S) - H(S|R)$. En aquest cas, $H(S) = \log(4)$ i $H(S|R) = \frac{9}{16}H(\frac{4}{9}, \frac{2}{9}, \frac{2}{9}, \frac{1}{9}) + \frac{3}{16}H(0, \frac{2}{3}, 0, \frac{1}{3}) + \frac{3}{16}H(0, 0, \frac{2}{3}, \frac{1}{3}) + \frac{1}{16}H(0, 0, 0, 1) = \frac{9}{16}1.83 + 2 \cdot \frac{3}{16}0.92 + 0 = 1.36$. Per tant, $I(S, R) = 2 - 1.365 = 0.635$.

2. (2.5 punts, 0.25+0.5+0.75+0.5+0.5) Volem comprimir el missatge m="SENEN TÉ SIS NENS I SET NENES", **prescindint dels accents ortogràfics i dels espais en blanc**:

   (a) A partir del missatge m, calculeu la freqüència i la probabilitat corresponent a cada caràcter, seguint l'ordre de la taula següent:

Símbol	Freqüència	Probabilitat
S		
E		
N		
T		
I		

(b) Suposant que una font S emet símbols amb aquestes probabilitats, quina és l'entropia de la font?

(c) Trobeu un codi **binari** òptim per a aquesta font.

(d) Quina és la longitud mitjana del codi que heu trobat?

(e) Quina és l'eficiència del codi que heu trobat?

Solució:

(a) Els símbols que apareixen en el missatge són:

Símbol	Freqüència	Probabilitat
S	6	6/23
E	7	7/23
N	6	6/23
T	2	2/23
I	2	2/23

(b) Per a calcular l'entropia de la font S, $H(S) = H(\frac{6}{23}, \frac{7}{23}, \frac{6}{23}, \frac{2}{23}, \frac{2}{23})$.

$$H(S) = \log 23 - \left\{ \frac{7}{23}\log 7 + \frac{6}{23}\log 6 + \frac{6}{23}\log 6 + \frac{2}{23}\log 2 + \frac{2}{23}\log 2 \right\}$$

$$= \log 23 - \left\{ \frac{7}{23}\log 7 + \frac{12}{23}\log 6 + \frac{4}{23}\log 2 \right\} = \log 23 - \left\{ \frac{7}{23}\log 7 + \frac{12}{23}\log 3 + \frac{12}{23}\log 2 + \frac{4}{23}\log 2 \right\}$$

$$= \log 23 - \left\{ \frac{7}{23}\log 7 + \frac{12}{23}\log 3 + \frac{16}{23}\log 2 \right\} = \log 23 - \left\{ \frac{7}{23}\log 7 + \frac{12}{23}\log 3 + \frac{16}{23}\log 2 \right\}$$

$$= 4.52 - \left\{ \frac{7}{23}\cdot 2.80 + \frac{12}{23}\cdot 1.58 + \frac{16}{23}\cdot 1 \right\} = 4.52 - \frac{1}{23}\{19.6 + 18.96 + 16\} = 4.52 - \left\{ \frac{54.56}{23} \right\}$$

$$= 4.52 - \{2.36\} = 2.16.$$

(c) Utilitzem el mètode de Huffman per trobar un codi instantani òptim.

Símbol	Freqüència	Probabilitat	Paraula-codi	Longitud
S	6	6/23	10	2
E	7	7/23	11	2
N	6	6/23	01	2
T	2	2/23	001	3
I	2	2/23	000	3

(d) La longitud mitjana és
$\overline{L} = 2 \cdot \frac{7}{23} + 2 \cdot \frac{6}{23} + 2 \cdot \frac{6}{23} + 3 \cdot \frac{2}{23} + 3 \cdot \frac{2}{23} = \frac{1}{23}(14 + 12 + 12 + 6 + 6) = \frac{1}{23}(50) = 2.17.$

(e) L'eficiència és $\eta = \frac{H(S)}{\overline{L}} = \frac{2.16}{2.17} = 0.995.$

3. (2.5 punts, 1+1+0.5) Volem comprimir el missatge m="SENEN TÉ SIS NENS I SET NENES", **prescindint dels accents ortogràfics i dels espais en blanc**. Comprimiu la cadena fent servir LZ77 amb la mida de diccionari D i la mida de buffer B següents:

   (a) $D = 18$ i $B = 3$.

   (b) $D = 3$ i $B = 3$.

   (c) Considereu que cada caràcter ocupa 8 bits i la mida del diccionari i del buffer determinen el nombre de bits de cada índex. Sabent que LZ77 necessita, respectivament, 11 i 15 tuples per a cadascun dels exemples anteriors, doneu (en forma de fracció) la taxa de compressió obtinguda en cada cas.

   Solució:

   (a+b) La codificació de la cadena amb les diferents mides del diccionari i del buffer són:

D=18, B=3		D=3, B=3	
(0,0,S)	S	(0,0,S)	S
(0,0,E)	E	(0,0,E)	E
(0,0,N)	N	(0,0,N)	N
(2,2,T)	ENT	(2,2,T)	ENT
(3,1,S)	ES	(3,1,S)	ES
(0,0,I)	I	(0,0,I)	I
(2,1,N)	SN	(2,1,N)	SN
(8,2,S)	ENS	(0,0,E)	E
(6,2,E)	ISE	(2,1,S)	NS
(12,1,N)	TN	(0,0,I)	I
(18,3,S)	ENES	(2,1,E)	SE
		(0,0,T)	T
		(0,0,N)	N
		(3,1,N)	EN
		(2,1,S)	ES

   (c) La cadena inicial ocupa $8 \cdot 23 = 184$ bits.
   En el primer cas, necessitem 5 bits per codificar la posició en el diccionari i 2 bits per al buffer. Per tant, cada 3-tupla ocupa $5 + 2 + 8 = 15$ bits. En total, la cadena comprimida ocupa $11 \cdot 15 = 165$ bits i la taxa de compressió és $R = \frac{165}{184} = 0.896$ bpb.
   En el segon cas, necessitem 2 bits per codificar la posició en el diccionari i 2 bits per al buffer. Per tant, cada 3-tupla ocupa $2 + 2 + 8 = 12$ bits. En total, la cadena comprimida ocupa $15 \cdot 12 = 180$ bits i la taxa de compressió és $R = \frac{180}{184} = 0.978$ bpb.

4. (2.5 punts, 0.75+0.75+1) Sigui $\{A_1, A_2, A_3, A_4\}$ el conjunt d'entrades i $\{B_1, B_2, B_3\}$ el conjunt de sortides d'un canal, amb matriu de probabilitats condicionades:

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

   (a) - Quina distribució inicial fa que $H(B)$ sigui màxima?
   - I quina distribució inicial fa que $H(B)$ tingui el màxim valor que podem obtenir si sabem que $p(A_3)$ és $\frac{1}{2}$?

(b) - Per a quina distribució inicial tenim $I(A,B) = \log(3)$?
- I per a quina distribució inicial tenim $I(A,B) = \log(4)$?

(c) - Amb distribució inicial $\{\frac{1}{8}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}\}$, quina és la regla a mínima probabilitat d'error (MPE)?
- I la probabilitat mitjana d'error fent servir aquest regla?

Solució:

(a) Si la distribució de probabilitats inicial és $\{p_1, p_2, p_3, p_4\}$, aleshores la matriu de probabilitats conjuntes és la següent:

$$\begin{pmatrix} 0 & p_1 & 0 \\ p_2 & 0 & 0 \\ 0 & 0 & p_3 \\ 0 & p_4 & 0 \end{pmatrix}.$$

Aleshores, la distribució de probabilitat final és $\{p_2, p_1 + p_4, p_3\}$. Tenim que $H(B)$ és màxima, $\log(3)$ si la distribució final és equiprobable; és a dir si $p_2 = p_3 = \frac{1}{3}$ i $p_1 + p_4 = \frac{1}{3}$. Per tant, una distribució inicial que fa que $H(B) = \log(3)$ seria $\{\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6}\}$.

Si la distribució de probabilitat inicial és $\{p_1, p_2, \frac{1}{2}, p_4\}$, aleshores la distribució final és, $\{p_2, p_1 + p_4, \frac{1}{2}\}$. Aleshores, l'entropia de B és $H(B) = H(p_2, p_1 + p_4, \frac{1}{2})$. El valor màxim d'entropia de B que podem aconseguir és quan $p_2 = (p_1 + p_4) = \frac{1}{4}$. Per tant, per exemple, per a distribució inicial $\{\frac{1}{8}, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$.

(b) Es tracta d'un canal determinista i tenim que $H(B|A) = 0$, per tant, $I(A,B) = H(B) - H(B|A) = H(B)$. La capacitat del canal és $C = max_{\{p(B_j)\}} H(B) = \log(3) = 1.58$. Aquesta capacitat s'assoleix quan la distribució final és equiprobable. Per tant, una distribució inicial que fa que $I(A,B) = \log(3)$ seria $\{\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6}\}$.

Com que la capacitat del canal és $C = \log(3) = 1.58$, aleshores no és possible que $I(A,B) = \log(4) = 2$ per a cap distribució de probabilitats inicial.

(c) Si la distribució de probabilitat inicial és $\{\frac{1}{8}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}\}$ aleshores la matriu de probabilitats conjuntes és la següent:

$$\begin{pmatrix} 0 & \frac{1}{8} & 0 \\ \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{3}{8} \\ 0 & \frac{1}{4} & 0 \end{pmatrix}.$$

Fixant-nos en els valors màxims a cada columna obtenim la següent funció de descodificació a mínima probabilitat d'error:

$$\begin{aligned} B_1 &\longrightarrow A_2 \\ B_2 &\longrightarrow A_4 \\ B_3 &\longrightarrow A_3 \end{aligned}$$

Aleshores, la probabilitat mitjana d'error en la descodificació és

$$\overline{p}_e = 1 - \frac{1}{4} - \frac{1}{4} - \frac{3}{8} = 1 - \frac{7}{8} = \frac{1}{8} = 0.125.$$