Examen 4 de abril 2013, preguntas y respuestas

Informació i Seguretat (Universitat Autònoma de Barcelona)

INFORMACIÓ I SEGURETAT
4 d'abril de 2013

Nom i cognoms: ______________________________ Grup: ______

- Cal que justifiqueu convenientment totes les respostes
- Valoració dels exercicis: 1) 1+1+0,5 punts; 2) 0,5+1+1 punts; 3) 1,25+1,25 punts; 4) 1+1+0,5 punts
- $\log 3 = 1.58$, $\log 5 = 2.32$, $\log 7 = 2.8$

1. Tenim dues urnes, amb tres boles cadascuna, que contenen diferents premis. Les boles de la primera urna contenen els premis $\{0€, 10€, 20€\}$ i les de la segona els premis $\{10€, 20€, 30€\}$. Llancem una moneda perfecta. Si surt cara prenem una bola de la primera urna i si surt creu prenem una bola de la segona urna.

   (a) Justifiqueu fent servir teoria de la informació si obtenim més informació sobre el resultat del llançament de la moneda quan el premi final és 20€, o quan el premi és 30€. Calculeu la informació en cada cas.

   (b) Quina és la informació obtinguda sobre el resultat del llançament de la moneda si coneixem el premi obtingut?

   (c) Quina és la informació obtinguda sobre el resultat del llançament de la moneda segons el premi obtingut?

Solució: Definim $X = \{x_1, x_2\} = \{c, +\}$ i $Y = \{y_1, y_2, y_3, y_4\} = \{0€,\ 10\ €,\ 20€,\ 30€\}$.

Calculem les probabilitats condicionades $p(x_i|y_j)$, les probabilitats conjuntes $p(x_i, y_j)$ i les probabilitats condicionades $p(y_j|x_i)$, tenint en compte que $p(x_i) = \frac{1}{2}, \forall i$.

$p(y_i\|x_j)$	0	10	20	30
c	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0
$+$	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

$p(y_i, x_j)$	0	10	20	30
c	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	0
$+$	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$p(x_i\|y_j)$	0	10	20	30
c	1	$\frac{1}{2}$	$\frac{1}{2}$	0
$+$	0	$\frac{1}{2}$	$\frac{1}{2}$	1

(a) Quan el premi obtingut és 30€, sabem que el resultat del llançament de la moneda ha estat per força creu. Per tant, no hi ha informació; en aquest cas la informació és 0. Si el premi obtingut és 20€, aleshores el resultat del llançament de la moneda podria haver estat tant cara com creu i, per tant, la informació en aquest cas és superior. Fent servir els valors obtinguts a les taules anteriors, tenim:

$$H(X|Y=20) = p(c|20)\log\frac{1}{p(c|20)} + p(+|20)\log\frac{1}{p(+|20)} = \frac{1}{2}\log 2 + \frac{1}{2}\log 2 = 1 \text{ bit.}$$

$$H(X|Y=30) = p(c|30)\log\frac{1}{p(c|30)} + p(+|30)\log\frac{1}{p(+|30)} = 1\log 1 + 0\log 0 = 0 \text{ bits.}$$

(b) $H(X|Y) = \sum_{i=1}^{2}\sum_{j=1}^{4} p(x_i, y_j)\log\frac{1}{p(x_i|y_j)} = 2\frac{1}{6}\log 1 + 4\frac{1}{6}\log 2 = \frac{4}{6}$ bits.

(c) Volem calcular $I(X,Y) = H(X) - H(X|Y)$.
Com els resultats del llançament de la moneda són equiprobables, tenim que $H(X) = 1$ bit. Per tant, $I(X,Y) = 1 - \frac{4}{6} = \frac{2}{6}$ bits.

2. Considereu una font $S = \{a_1, a_2, a_3, a_4, a_5\}$ amb probabilitats $\{\frac{3}{10}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{10}\}$ i el codi $C = \{00, 10, 01, 000, 001\}$.

   (a) És C un codi de descodificació única? Per què?

   (b) Determineu si l'eficiència de C és igual o és menor que 1.

   (c) És un codi òptim? Per què?

   Solució:

   (a) Tenim que C no és de descodificació única ja que, per exemple la seqüència 000000 podria ser tant $a_1 a_1 a_1$ com $a_4 a_4$.

   (b) Sabem que $\eta = \frac{H(S)}{\bar{L}\log D}$, on $D = 2$.
   Tenim que $H(S) = \frac{3}{10}\log\frac{10}{3} + \frac{3}{5}\log 5 + \frac{1}{10}\log 10 = \frac{3}{10}(\log 2 + \log 5 - \log 3) + \frac{3}{5}\log 5 + \frac{1}{10}(\log 2 + \log 5) = \frac{4}{10}\log 2 + \log 5 - \frac{3}{10}\log 3 = 0.4 + 2.3219 - \frac{3}{10}1.5850 = 2.2464$ bits.
   Per una altra banda, $\bar{L} = \frac{6}{10} + 2\frac{2}{5} + \frac{3}{5} + \frac{3}{10} = \frac{23}{10} = 2.3$.
   Aplicant la fòrmula anterior tenim que $\eta = \frac{2.2464}{2.3}$ és menor que 1.

   (c) Per saber si el codi és òptim, busquem un codi de Huffman. Aplicant l'algorisme obtenim $H = \{10, 00, 01, 110, 111\}$ que té longitud mitjana $\bar{L}_H = 2.3$. El codi anterior té la mateixa longitud mitjana que el codi de Huffman. Tot i això, com el codi no és de descodificació única no és un codi òptim.

3. Compressió de texts.

   (a) Comprimiu el següent mapa de bits fent servir la tècnica més addient i digueu quina és la taxa de compressió.

   11111111111111111
   11111111111111110
   00000000001111100
   00000000011111000
   00000000111110000
   00000001111100000
   00000011111000000
   00000111110000000
   00001111100000000
   00011111000000000
   00111110000000000

   (b) Quin mètode dóna una millor taxa de compressió de la cadena "THIS IS THE HOUSE OF THE MOUSE.", l'algorisme LZ77 o LZ78? Doneu tots el passos, considereu $|D| = 13$,$|B| = 4$ i que cada enter ocupa 4 bits i cada caràcter 8 bits.

   Solució:

   (a) En aquest cas, aplicarem l'algorisme RLE. Considerem que tenim una matriu 11×17 de bits. La longitud de cada fila es pot emmegatzemar en 5 bits. La mida de la imatge seria $11 \times 17 + 5 = 192$ bits.
   La codificaició RLE seria,

Fila	*Codificació*	*Fila*	*Codificació*
1	0 17	7	6 5 6
2	0 16 1	8	5 5 7
3	10 5 2	9	4 5 8
4	9 5 3	10	3 5 9
5	8 5 4	11	2 5 10
6	7 5 5		

Si cada valor s'emmagatzema en 5 bits, necessitarem $32 \times 5 + 5 = 165$ bits. Per tant, $R = \frac{165}{192}$ bpb.

(b) Fent servir LZ77 tenim la següent compressió: $(0, 0, T), (0, 0, H), (0, 0, I), (0, 0, S), (0, 0, ␣)$, $(3, 3, T)(8, 1, E), (4, 1, H), (0, 0, O), (0, 0, U), (9, 1, E), (6, 1, O), (0, 0, F), (13, 4, ␣), (0, 0, M)$, $(13, 4, .)$.

La taula obtinguda amb l'algorisme LZ78 és:

	Dicc	*Codi*		*Dicc*	*Codi*
0	null				
1	T	(0,T)	11	U	(0,U)
2	H	(0,H)	12	SE	(4,E)
3	I	(0,I)	13	␣O	(5,O)
4	S	(0,S)	14	F	(0,F)
5	␣	(0,␣)	15	␣TH	(7,H)
6	IS	(3,S)	16	E	(0,E)
7	␣T	(5,T)	17	␣M	(5,M)
8	HE	(2,E)	18	OU	(10,U)
9	␣H	(5,H)	19	SE.	(12,.)
10	O	(0,O)			

Si considerem que codifiquem cada caràcter amb 8 bits i cada índex enter amb 4 bits, aleshores el text original ocupa $31 * 8 = 248$ bits. La taxa de compressió fent servir LZ77 és $R = \frac{16(4+4+8)}{248} = \frac{256}{248}$. En el cas de LZ78, tenim una taxa de compressió de $\frac{19(4+8)}{248} = \frac{228}{248}$ i per tant, la taxa de compressió és menor en aquest cas.

4. Considereu el canal determinat per la matriu de transicions:

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

(a) Doneu la informació mútua de la entrada i la sortida, si la distribució inicial de probabilitats és $(\frac{1}{3}, \frac{1}{3}, 0, \frac{1}{3})$.

(b) És possible tenir una informació mútua de 1.9 bits? En cas positiu digueu per a quina distribució inicial tenim aquesta informació mútua i en cas negatiu raoneu per què no podem.

(c) Si la distribució de probabilitats inicial és $(\frac{1}{3}, \frac{1}{3}, 0, \frac{1}{3})$, calculeu la regla MPE i digueu quina és la probabilitat mitjana d'error.

Solució:

(a) $I(A, B) = H(A) - H(A|B)$.

$p(b_i, a_j)$	b_1	b_2	b_3	b_4
a_1	0	$\frac{1}{12}$	0	$\frac{1}{4}$
a_2	$\frac{1}{12}$	0	$\frac{1}{4}$	0
a_3	0	0	0	0
a_4	0	$\frac{1}{4}$	0	$\frac{1}{12}$

$p(a_j \vert b_i)$	b_1	b_2	b_3	b_4
a_1	0	$\frac{1}{4}$	0	$\frac{3}{4}$
a_2	1	0	1	0
a_3	0	0	0	0
a_4	0	$\frac{3}{4}$	0	$\frac{1}{4}$

Tenim que $H(A) = H(\frac{1}{3}, \frac{1}{3}, 0, \frac{1}{3}) = \log 3 = 1.58$. D'altra banda, $H(A|B) = 2\frac{1}{12}\log 4 + 2\frac{1}{4}\log\frac{4}{3} = \frac{1}{3} + \frac{1}{2}(\log 4 - \log 3) = \frac{1}{3} + \frac{1}{2}(0.42) = 0.33 + 0.21 = 0.54$. Finalment, $I(A, B) = 1.58 - 0.54 = 1.04$ bits.

(b) En aquest canal, tenim $C = \log 4 - H$, on $H = H(0, \frac{1}{4}, 0, \frac{3}{4}) = \frac{1}{4}\log 4 + \frac{3}{4}\log\frac{4}{3} = \frac{1}{2} + \frac{3}{4}(2 - 1.58) = 0.35 + 0.31 = 0.81$. Per tant, $C = 2 - 0.81 = 1,19$ bits. Així, la informació màxima del canal és 1.19 i no pot haver-hi cap distribució inicial que faci que la informació mútua sigui 1.9.

(c) La Regla MPE és aquella que assigna a cada b_j el valor a_i tal que $p(a_i|b_j)$ sigui màxima. Mirant les taules de l'apartat a) tenim $f(b_1) = a_2, f(b_2) = a_4, f(b_3) = a_2, f(b_4) = a_1$. Aleshores la probabilitat mitjana d'error és $1 - (\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{12}) = 1 - \frac{10}{12} = \frac{1}{6}$.