Semestre : 6

Unité d'enseignement : UEF 3.2.2

Cours 4

Codage et Théorie de l'information

Dr Mahmoud Hadef

Codage de l'information

Définition du Code Efficace

Formellement, si l'on rappelle que le symbole-source $u_i\ (1 \leq i \leq N)$ a une probabilité p_i d'être émis, et si l'on dénote l_i la longueur du mot de code correspondant, *l'espérance de la longueur de code* $*$ Si nous prenons $p_1 = p_5 = 0.1, p_2 = p_4 = 0.2\ et\ p_3 = 0.4$

l'espérance de la longueur de code $*\ E[L]$ est l'espérance de la longueur d'un mot de code, c'est-à-dire

$$\boldsymbol{E[L] = \sum_{i=1}^{N} p_i l_i}$$

Codage de l'information

Définition de l'Arbre n-aire probabilisé

Un arbre n-aire probabilisé est un arbre n-aire ayant des nombres positifs entre 0 et 1 («probabilités») affectés à chaque nœud (y compris les feuilles) de sorte que :

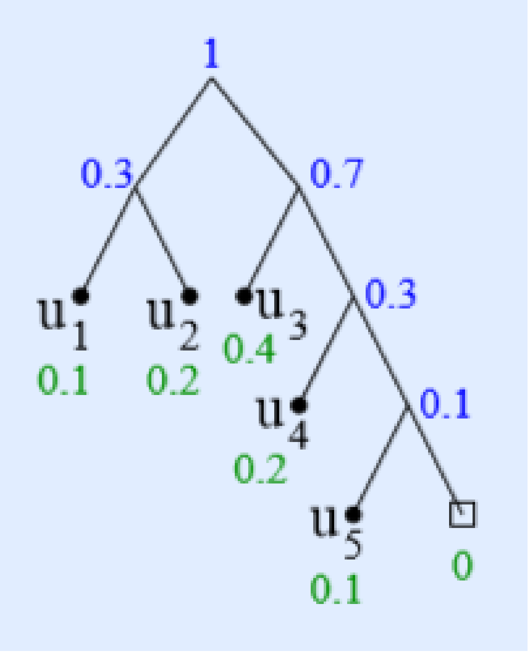
1. La racine est dotée d'une probabilité 1,

2. La probabilité de chaque nœud (y compris la racine) est la somme des probabilités de ses fils.

Codage de l’information

Exemple 10

Si nous prenons $p_1 = p_5 = 0.1, p_2 = p_4 = 0.2\ et\ p_3 = 0.4$ pour le code binaire instantané de l’exemple précèdent, nous obtiendrons l’arbre binaire suivant, avec les probabilités :



Codage de l'information

Définition du Lemme Longueur du Chemin

Dans un arbre n-aire probabilisé, la profondeur moyenne des feuilles est égale à la somme des probabilités des nœuds intérieurs (c'est-à-dire pas les feuilles, mais racine comprise).

Codage de l’information

Exemple 11:

du Lemme Longueur du Chemin

Dans l’exemple précédent, la profondeur moyenne des feuilles était : $1 + 0.3 + 0.7 + 0.3 + 0.1 = 2.4$ par le lemme de la longueur de chemin. À titre de vérification, notez que (définition de l’espérance de la longueur de code) $2 \cdot 0.1 + 2 \cdot 0.2 + 2 \cdot 0.4 + 3 \cdot 0.2 + 4 \cdot 0.1 = 2.4$

Codage de l’information

Définition de L’Entropie d’une feuille d’arbre n-aire probabilisé

Définition: Soit N le nombre de feuilles d’un arbre n-aire probabilisé et $p_1 , p_2 \, p_3 \;\; , p_N$ leurs probabilités. L’entropie des feuilles d’un tel arbre est définie comme:

$$\boldsymbol{H_{feuilles}} = -\sum_{\boldsymbol{i}} \boldsymbol{p_i log\, p_i}$$

Propriété: Pour l’arbre n-aire probabilisé correspondant à l’arbre de codage instantané d’une source d’information *U*, nous avons : $\boldsymbol{H_{feuilles}} = \boldsymbol{H(U)}$

Codage de l'information

Exemple 12

Supposons que les $M = 5$ nœuds pour l'arbre des exemples 9 and 10 et , soient numérotés de telle façon que

$$P_1 = 1, P_2 = 0.3, P_3 = 0.7, P_4 = 0.3\ et P_5 = 0.1$$

Alors $H_{feuilles} = -\sum_{i=1}^{5} p_i log p_i \cong 2.12\ bit$.

Nous avons $n_1 = 2\ et\ q_{11} = 0.3\ et\ q_{12} = 0.7$, ainsi

$$H_1 = -0.3 log 0.3 - 0.7 log 0.7 \cong 0.881 bit.$$

De même, $n_2 = 2\ et\ q_{21} = 0.1\ et\ q_{22} = 0.2$, ainsi

$$H_2 = -\frac{0.1}{0.3} log \frac{0.1}{0.3} - \frac{0.2}{0.3} log \frac{0.2}{0.3} \cong 0.918 bit.$$

Montrez que $H_3 \cong 0.985\ bit, H_4 \cong 0.918\ bit, H_5 \cong 0.$

Codage de l'information

Théorème de l'entropie des feuilles

L'entropie des feuilles d'un arbre n-aire probabilisé est égale à la somme sur tous les nœuds intérieurs (racine comprise) de l'entropie de branchement de ce nœud pondérée par sa probabilité. En employant les notations définies ci-dessus :

$$H_{feuilles} = \sum_{i=1}^{M} P_i H_i$$

Codage de l’information

Exemple 11

Nous calculons $\boldsymbol{H_{feuille}}$ pour obtenir

$$\boldsymbol{H_{feuille} = 1.H_1 + 0.3H_2 + 0.7H_3 + 0.3\,H_4 + 0.1H_5}$$

$$\boldsymbol{H_{feuille} \cong .881 + 0.3 * 0.918 + 0.7 * 0.985 + 0.3 * 0.918 + 0\;bit \cong 2.122\;bit}$$

conformément au calcul direct effectué dans l’exemple précèdent.

Codage de l’information

Théorème

Pour deux codes instantanés de la même source d’information, celui qui a la longueur moyenne du code la plus courte a le plus haut taux d’entropie par symbole.

Codage de l'information

Théorème de Shannon sur le codage, 1ère partie

Pour toute source d'information discrète sans mémoire d'entropie $H(U)$, l'espérance de la longueur de code $E[L]$ de tout code D-aire instantané pour cette source satisfait :

$$E[L] \geq \frac{H(U)}{\log D}$$

Codage de l’information

Théorème de Shannon sur le codage, 2ème partie

Pour toute source d’information discrète sans mémoire d’entropie $H(U)$, il existe au moins un code instantané D-aire dont l’espérance de la longueur de code $E[L]$ satisfait:

$$E[L] \geq \frac{H(U)}{\log D} + 1$$

Codage de l'information

Exemple 12

Considérons le codage binaire $\mathrm{D} = 2$ de Shannon-Fano pour la source d'arité $4\,U$ pour laquelle $p_1 = 0.4, p_2 = 0.3, p_3 = 0.2, p_4 = 0.1$

Un tel codage aura pour longueurs de mots de code (étant donné que $log_2(2) = 1$)

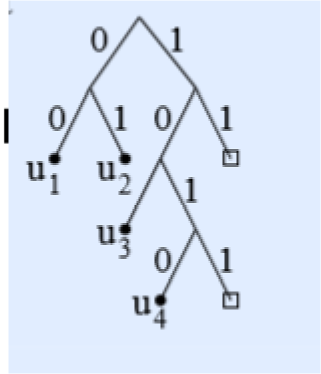
$$l_1 = -\lceil log_2 0.4 \rceil = 2 \qquad l_2 = -\lceil log_2 0.3 \rceil = 2$$

$$l_3 = -\lceil log_2 0.2 \rceil = 3 \qquad l_2 = -\lceil log_2 0.1 \rceil = 4$$

Nous construisons donc le code par l'algorithme donné dans la démonstration de l'inégalité de Kraft, pour obtenir le code dont l'arbre binaire est

Codage de l'information

Exemple 12



Par le lemme de la longueur de chemin, nous avon

$$E[L] = 1 + 0.7 + 0.3 + 0.3 + 0.1 = 2.4$$

et un calcul direct donne:

$$H(U) = 0.4\,log\,0.4 + 0.3\,log\,0.3 + 0.2\,log\,0.2 + 0.1 log 0.1 \quad \cong \textit{1.8 bit}$$

Notez toutefois que ce code est clairement non-optimal. Si nous avions simplement employé les 4 mots de code possibles de longueur 2, nous aurions eu un code plus court ($E[L] = 2$)

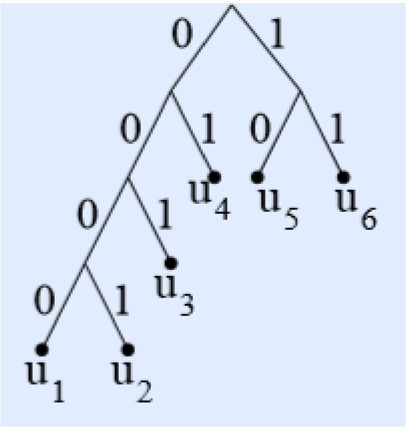
Codage de l'information

Exemple 13: Codage binaire de Huffman

Considérons une source d'information telle que

U	u_1	u_2	u_3	u_4	u_5	u_6
p_i	0.05	0.1	0.15	0.27	0.20	0.23

Un code de Huffman pour U est donné par :



z_1	z_2	z_3	z_4	z_5	z_6
0000	0001	001	01	10	11

Codage de l'information

Exemple 13: Codage binaire de Huffman

Les probabilités associées aux nœuds intérieurs sont les suivantes :

$v_1 = u_1 \oplus u_2$	$v_2 = v_1 \oplus u_3$	$v_3 = u_5 \oplus u_6$	$v_4 = v_2 \oplus u_4$	$v_5 = v_4 \oplus v_3$
0.15	0.30	0.43	0.57	1

Enfin, notez que

$E[L] = 2(0.2 + 0.23 + 0.27) + 3(0.15) + 4(0.1 + 0.05) = 2.45$

(ou par le lemme de la longueur de chemin:

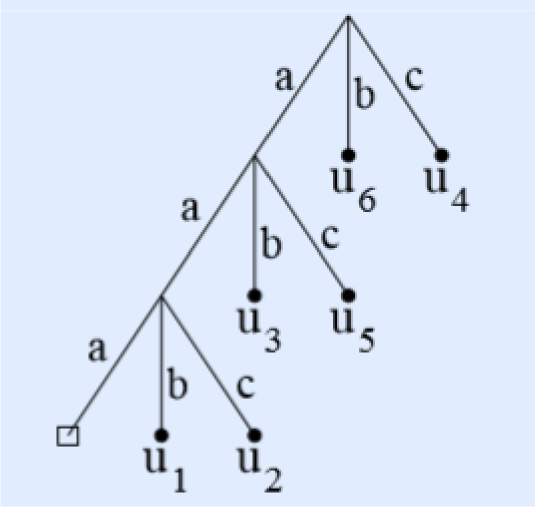
$E[L] = 1 + 0.57 + 0.43 + 0.30 + 0.15 = 2.45$), et

$$\mathrm{H}[U] = -\sum_{i=1}^{6} p_i \, log p_i = 2.42 \, bit.$$

Codage de l'information

Exemple 14: Codage binaire de Huffman

Pour la même source U de l'exemple précédent et en employant un code ternaire $(D = 3)$, nous avons pour le reste de $1 - n \coloneqq 1 - 6 = -5$ par $D - 1 \coloneqq 2$: $r = 1$. En effet, -5 $= -3 \cdot 2 + 1$. Donc *une* feuille inutilisée doit être introduite. Le code ternaire de Huffman est dans ce cas:



z_1	z_2	z_3	z_4	z_5	z_6
aab	aac	ab	c	ac	b

Codage de l'information

Exemple 14: Codage binaire de Huffman

Les probabilités associées à des nœuds intérieurs sont :

$v_1 = u_1 \oplus u_2$	$v_2 = v_1 \oplus u_3 \oplus u_5$	$v_3 = v_2 \oplus u_6 \oplus u_4$
0.15	0.50	1

Enfin notez que

$E[L] = 1 + 0.5 + 0.15 = 1.65$

$(par\ le\ lemme\ de\ la\ longeur\ de\ chemin) et$

$$\frac{H(U)}{log3} = \frac{2.42}{1.59} = 1.52$$

Codage de l’information

Codage de Shannon-Fano

<table>
<tr><th>Symboles
s_k</th><th>Proba
$p(s_k)$</th><th></th><th></th><th></th><th></th><th>Mots-codes
c_k</th><th>Longueur
l_k</th></tr>
<tr><td>s_1</td><td>0.25</td><td rowspan="2">0</td><td>0</td><td rowspan="2"></td><td rowspan="2"></td><td>00</td><td>2</td></tr>
<tr><td>s_2</td><td>0.25</td><td>1</td><td>01</td><td>2</td></tr>
<tr><td>s_3</td><td>0.125</td><td rowspan="6">1</td><td rowspan="2">0</td><td>0</td><td rowspan="2"></td><td>100</td><td>3</td></tr>
<tr><td>s_4</td><td>0.125</td><td>1</td><td>101</td><td>3</td></tr>
<tr><td>s_5</td><td>0.0625</td><td rowspan="4">1</td><td rowspan="2">0</td><td>0</td><td>1100</td><td>4</td></tr>
<tr><td>s_6</td><td>0.0625</td><td>1</td><td>1101</td><td>4</td></tr>
<tr><td>s_7</td><td>0.0625</td><td rowspan="2">1</td><td>0</td><td>1110</td><td>4</td></tr>
<tr><td>s_8</td><td>0.0625</td><td>1</td><td>1111</td><td>4</td></tr>
</table>

Codage de l'information

Codage de Shannon-Fano

<table>
<tr><th>a_i</th><th>$p(a_i)$</th><th>1</th><th>2</th><th>3</th><th>4</th><th>Code</th></tr>
<tr><td>a_1</td><td>0.36</td><td rowspan="2">0</td><td colspan="3">00</td><td>00</td></tr>
<tr><td>a_2</td><td>0.18</td><td colspan="3">01</td><td>01</td></tr>
<tr><td>a_3</td><td>0.18</td><td rowspan="4">1</td><td colspan="3">10</td><td>10</td></tr>
<tr><td>a_4</td><td>0.12</td><td rowspan="3">11</td><td colspan="2">110</td><td>110</td></tr>
<tr><td>a_5</td><td>0.09</td><td rowspan="2">111</td><td>1110</td><td>1110</td></tr>
<tr><td>a_6</td><td>0.07</td><td>1111</td><td>1111</td></tr>
</table>

Codage de l'information

Codage de Huffman

s_k $[R_0]$	$p(s_k)$	c_k	$[R_1]$		$[R_2]$		$[R_3]$		$[R_4]$	
									0,60	(0)
							0,40	(1)	0,40	(1)
s_1	0,30	00	0,30	(00)	0,30	(00)	0,30	(00)		
					0,30	(01)	0,30	(01)		
s_2	0,25	10	0,25	(10)	0,25	(10)				
s_3	0,15	11	0,15	(11)	0,15	(11)				
s_4	0,15	010	0,15	(010)						
			0,15	(011)						
s_5	0,10	0110								
s_6	0,05	0111								

Codage de l’information

Codage de Huffman

j'aime aller sur le bord de l'eau les jeudis ou les jours impairs

`	m	j	o	d	a	i	r	u	l	s	e	
2	2	3	3	3	4	4	5	5	6	6	8	12

Références

- F. Bavaud , J.-C. Chappelier , J. Kohlas ; Introduction à la Théorie de l'Information et ses applications ; Université de Fribourg.
- O. Rioul ; Théorie de l’information et du codage ; Lavoisier, 2007.
- Y. Mori ; Théorie de l'information et du codage : signal analogique, signal numérique et applications en télécommunications ; Hermès Science, 2006.
- T. M. Cover and J. A. Thomas; Elements of information theory, 2nd edition, Wiley Series in telecommunications and signal Processing, 2006.