Compression de données

Juin 2018

Durée totale : 1h30
Les documents/calculatrices sont autorisés.

Exercice 1 - Compression conservative

On considère un fichier dont le texte est `CHOCHICHON⋆CHO⋆DE⋆CHIEN` défini sur l'alphabet $\Sigma = \{\star\texttt{CDEHINO}\}$. Le fichier F se termine par un caractère de fin dont on ne tiendra pas compte. On supposera que la distribution de probabilité des symboles est donnée par leur fréquence relative d'apparition dans le fichier.

1.1 Rappelez la définition de l'entropie et son interprétation au sens de la compression de données. Quelle est l'entropie du fichier F ?

Cf cours.

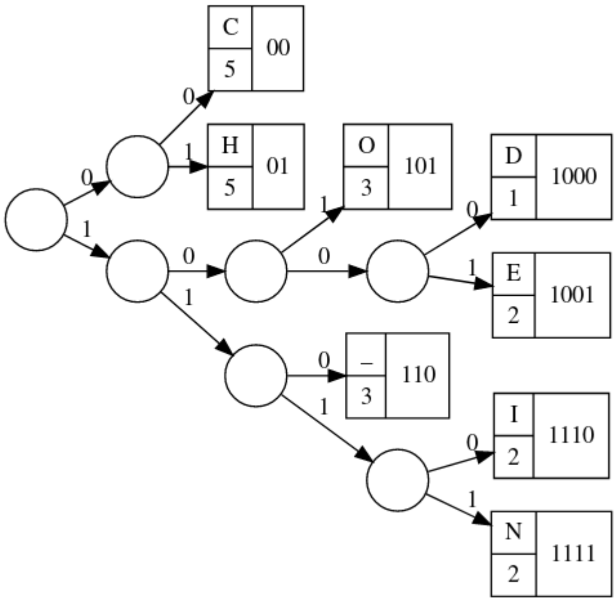
X	p(X)
D	0.043478
I	0.086957
H	0.217391
O	0.130435
C	0.217391
_	0.130435
N	0.086957
E	0.086957

Entropy = 2.839689958091771

1.2 Combien faut-il de bits pour représenter les symboles du fichier Quelle est la taille de F sans compression ?

L'alphabet fait 8 caractères donc 3 bits. Pour stocker le fichier, il faut donc $23 \times 3 = 69$ bits.

1.3 Calculez l'encodage de Huffman des symboles contenus dans F, en faisant clairement apparaître l'arbre de Huffman.



1.4 Quelle est la taille de F après compression par l'encodage de Huffman ? Comparez à la taille minimale prévue par l'entropie.

Après compression, le fichier fait 66 bits. La taille min d'entropie d'ordre 1 est $23 \times 2.84 = 65.32$ bits

On souhaite maintenant compresser le fichier F par l'algorithme LZW. *On considère que le dictionnaire par défaut $\mathcal{D}$ du compresseur ne contient que les symboles de Σ initialement (i.e. un tableau de 8 éléments avec les codes <0>...<7>)*

1.6 Compressez le fichier F avec l'algorithme LZW (incluez le tableau de compression dans votre copie).

```
{’D’: 0, ’I’: 1, ’H’: 2, ’O’: 3, ’C’: 4, ’_’: 5, ’N’: 6, ’E’: 7}
|w       |c       |w.c     |Ouput           |Dictionnaire    |
|--------|--------|--------|----------------|----------------|
|        |C       |C       |                |                |
|C       |H       |CH      |<4> (C)         |CH = <8>        |
|H       |O       |HO      |<2> (H)         |HO = <9>        |
|O       |C       |OC      |<3> (O)         |OC = <10>       |
|C       |H       |CH      |                |                |
|CH      |I       |CHI     |<8> (CH)        |CHI = <11>      |
|I       |C       |IC      |<1> (I)         |IC = <12>       |
|C       |H       |CH      |                |                |
|CH      |O       |CHO     |<8> (CH)        |CHO = <13>      |
|O       |N       |ON      |<3> (O)         |ON = <14>       |
|N       |_       |N_      |<6> (N)         |N_ = <15>       |
|_       |C       |_C      |<5> (_)         |_C = <16>       |
|C       |H       |CH      |                |                |
|CH      |O       |CHO     |                |                |
|CHO     |_       |CHO_    |<13> (CHO)      |CHO_ = <17>     |
|_       |D       |_D      |<5> (_)         |_D = <18>       |
|D       |E       |DE      |<0> (D)         |DE = <19>       |
|E       |_       |E_      |<7> (E)         |E_ = <20>       |
|_       |C       |_C      |                |                |
|_C      |H       |_CH     |<16> (_C)       |_CH = <21>      |
```

```
|H        |I        |HI       |<2> (H)            |HI = <22>        |
|I        |E        |IE       |<1> (I)            |IE = <23>        |
|E        |N        |EN       |<7> (E)            |EN = <24>        |
|N        |         |         |<6> (N)            |                 |
Final: CHO<8>I<8>ON_<13>_DE<16>HIEN
```

1.7 Quelle est la taille de F après compression LZW en supposant que les codes du dictionnaire utilisent un nombre de bits constant ?

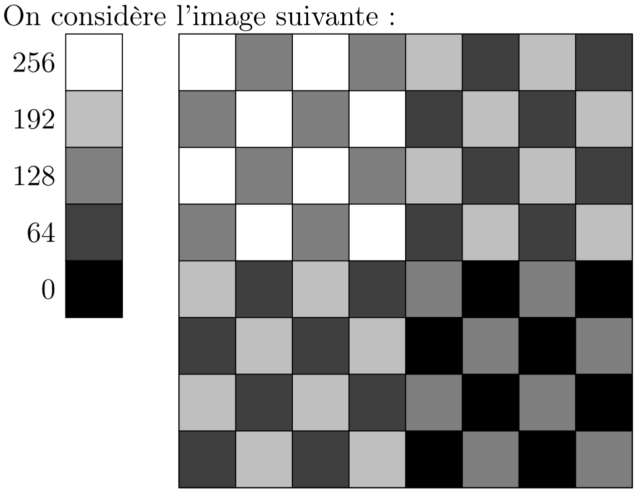
> En supposant, un nombre constant de bits d'encodage, il faut 5 bits/ch donc 5 * 18 = 90 bits.

1.8 Même question avec des codes de taille variable. Expliquez le résultat.

> Le dictionnaire initial fait 8 charactères, on peut démarrer initialement avec 8+1 codes (il faut un caractère pour signaler le passage à la quantification supérieure) soit 4 bits.
> La chaîne est alors CHO<9>I<9>ON_<14>_DE<OPCODE><17>HIEN soit $4 \times 14 + 5 \times 5 = 81$ bits.

Exercice 2 - Compression avec perte

On considère l'image suivante :



3.1 En supposant que l'image à des valeurs de niveau de gris, écrire la matrice correspondante à l'image.

3.2 Si on effectue un sous-échantillonage de cette image en prenant un pixel sur deux (interpolation au plus proche), quelle est l'image résultante ? Comment s'appelle ce phénomène et à quoi est-il dû d'un point de vue "théorie du signal" ?

3.3 Quelle est la transformée d'Hadamard de l'image d'entrée ? Choisir parmi les trois matrices et expliquer votre choix (soit par des calculs, soit par la logique).

$$\mathcal{Q}_1 = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix} \quad \mathcal{Q}_2 = \frac{1}{2\sqrt{2}} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}$$

$$\mathcal{Q}_3 = \frac{1}{8}\begin{bmatrix} 1024 & 256 & 0 & 0 & 0 & 0 & 0 & 0 \\ 256 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 512 \end{bmatrix}$$

3.4 Combien faut-il de bits pour encoder la matrice précédente sans compression ?

3.5 Proposer une méthode pour la compresser, et indiquer le gain obtenu avec votre méthode.

3.6 Rappeler et expliquer succintement les 4 étapes majeures de la compression JPEG d'une image en niveaux de gris.

3.7 En quoi l'étape de DCT diffère-t-elle de la tranformée d'Hadamard et pourquoi la prèfère-t-on ?