Théorie de l'Information (4)

Philippe Duchon

ENSEIRB

2007-08

Codage de Huffman (1)

Lemme (Huffman)

Pour toute source S, émettant les symboles $s_1, \ldots, s_N$ avec les probabilités $p_1 \geq p_2 \geq \cdots \geq p_N$, il existe un code binaire *optimal* d'assortiment de longueurs $\ell_1, \ldots, \ell_N$, vérifiant

1. $\ell_1 \leq \ell_2 \leq \cdots \leq \ell_N$
2. $\ell_{N-1} = \ell_N$
3. $C(s_{N-1})$ et $C(s_N)$ ont un préfixe commun de longueur $\ell_N - 1$.

Codage de Huffman (1)

Lemme (Huffman)

Pour toute source S, émettant les symboles $s_1, \ldots, s_N$ avec les probabilités $p_1 \geq p_2 \geq \cdots \geq p_N$, il existe un code binaire *optimal* d'assortiment de longueurs $\ell_1, \ldots, \ell_N$, vérifiant

1. $\ell_1 \leq \ell_2 \leq \cdots \leq \ell_N$
2. $\ell_{N-1} = \ell_N$
3. $C(s_{N-1})$ et $C(s_N)$ ont un préfixe commun de longueur $\ell_N - 1$.

(1) Si $\ell_{i+1} < \ell_i$, on peut échanger $C(s_i)$ et $C(s_{i+1})$ sans augmenter la longueur moyenne (cela implique que $p_i = p_{i+1}$, sinon violation de l'optimalité) ; on peut donc diminuer le nombre de décroissance de la suite des ℓ_i.

Codage de Huffman (1)

Lemme (Huffman)

Pour toute source S, émettant les symboles $s_1, \ldots, s_N$ avec les probabilités $p_1 \geq p_2 \geq \cdots \geq p_N$, il existe un code binaire *optimal* d'assortiment de longueurs $\ell_1, \ldots, \ell_N$, vérifiant

1. $\ell_1 \leq \ell_2 \leq \cdots \leq \ell_N$
2. $\ell_{N-1} = \ell_N$
3. $C(s_{N-1})$ et $C(s_N)$ ont un préfixe commun de longueur $\ell_N - 1$.

(2) Par la construction de la preuve du théorème de Kraft : lors de la construction du code pour $(s_1, \ldots, s_{N-1})$, il reste forcément un mot de longueur ℓ_{N-1} qui n'a pas de préfixe dans le code, et on peut prendre un tel mot pour coder s_N

Codage de Huffman (1)

Lemme (Huffman)

Pour toute source S, émettant les symboles $s_1, \ldots, s_N$ avec les probabilités $p_1 \geq p_2 \geq \cdots \geq p_N$, il existe un code binaire *optimal* d'assortiment de longueurs $\ell_1, \ldots, \ell_N$, vérifiant

1. $\ell_1 \leq \ell_2 \leq \cdots \leq \ell_N$
2. $\ell_{N-1} = \ell_N$
3. $C(s_{N-1})$ et $C(s_N)$ ont un préfixe commun de longueur $\ell_N - 1$.

Soit C un code optimal satisfaisant (1) et (2), et soit w le mot de longueur ℓ_N qui diffère de $C(s_{N-1})$ par son dernier bit.

- Si $w = C(s_N)$, C satisfait (3).
- Si $w = C(s_j)$ avec $j < N-1$, on peut échanger $C(s_j)$ et $C(s_N)$ et obtenir un code C' qui satisfait les 3 conditions.
- Sinon, on peut remplacer $C(s_N)$ par w

Codage de Huffman (1)

Lemme (Huffman)

Pour toute source S, émettant les symboles $s_1, \ldots, s_N$ avec les probabilités $p_1 \geq p_2 \geq \cdots \geq p_N$, il existe un code binaire *optimal* d'assortiment de longueurs $\ell_1, \ldots, \ell_N$, vérifiant

1. $\ell_1 \leq \ell_2 \leq \cdots \leq \ell_N$
2. $\ell_{N-1} = \ell_N$
3. $C(s_{N-1})$ et $C(s_N)$ ont un préfixe commun de longueur $\ell_N - 1$.

Remarque : pour certaines sources, l'assortiment de longueurs des codes optimaux peut ne pas être unique ; exemple : probabilités $(0.6, 0.15, 0.1, 0.05, 0.05)$ admet $\mathcal{L} = (1, 2, 3, 4, 4)$ ou $\mathcal{L}' = (1, 3, 3, 3, 3)$ qui ont tous deux longueur moyenne 1.8.

Codage de Huffman (2)

Pour la **construction** de codes optimaux :

Théorème de Huffman

Soit S une source émettant les symboles $s_1, \ldots, s_N$ avec probabilités $p_1 \geq \cdots \geq p_N$, et S^* une source émettant les symboles $s_1, \ldots, s_{N-2}, s^*_{N-1}$ avec probabilités $p_1, \ldots, p_{N-2}, p^*_{N-1} = p_{N-1} + p_N$. Alors

- Si C est un code optimal pour S, satisfaisant les conditions du lemme de Huffman, le code C^* obtenu en remplaçant les deux mots codant s_{N-1} et s_N par leur préfixe commun (qui code s^*_{N-1}) est optimal pour S^*.
- Réciproquement, si C^* est un code optimal pour S^*, le code obtenu en remplaçant le mot $C^*(s^*_{N-1})$ par ses deux prolongements possibles, codant s_{N-1} et s_N, est optimal pour S.

Preuve du théorème de Huffman

On remarque tout d'abord que les longueurs moyennes $L(C)$ et $L(C^*)$ sont liées par

$$L(C) = L(C^*) + p_{N-1} + p_N.$$

Preuve du théorème de Huffman

On remarque tout d'abord que les longueurs moyennes $L(C)$ et $L(C^*)$ sont liées par

$$L(C) = L(C^*) + p_{N-1} + p_N.$$

Partie directe : supposons que le code C^* ne soit pas optimal pour S^*, il existerait alors un code C'^* de longueur strictement inférieure. Alors en appliquant la transformation $C^* \mapsto C$ à C'^* on obtiendrait un code pour S, de longueur moyenne strictement inférieure à celle de C, ce qui contredit l'optimalité de C.

Preuve du théorème de Huffman

On remarque tout d'abord que les longueurs moyennes $L(C)$ et $L(C^*)$ sont liées par

$$L(C) = L(C^*) + p_{N-1} + p_N.$$

Partie réciproque : supposons que le code C ne soit pas optimal pour S. Alors il existerait un code C' de longueur strictement inférieure. Le lemme de Huffman permet d'obtenir un code C'', de même longueur moyenne, et la transformation $C \mapsto C^*$ appliquée à C'' donne alors un code pour S^* de longueur strictement inférieure à celle de C^*, ce qui contredit l'optimalité de C^*.

Construction de codes de Huffman

- En appliquant récursivement la partie directe du théorème de Huffman (jusqu'à ne plus avoir qu'une source à 2 symboles, codée optimalement avec longueur 1), on construit (en partant des feuilles) un arbre (binaire) d'appariements des symboles de la source.
- La partie réciproque du théorème nous assure (encore récursivement) que, dans l'arbre obtenu, la profondeur de chaque feuille (singleton) nous donne la longueur de son codage dans un code optimal.

Un exemple de codage de Huffman

$$S = ((a, 0.34)(b, 0.30)(c, 0.26)(d, 0.08)(e, 0.02))$$

$$(0.34, 0.30, 0.26, 0.08, 0.02)$$

Un exemple de codage de Huffman

$$S = ((a, 0.34)(b, 0.30)(c, 0.26)(d, 0.08)(e, 0.02))$$

$$(0.34, 0.30, 0.26, 0.08, 0.02)$$

Un exemple de codage de Huffman

$$S = ((a, 0.34)(b, 0.30)(c, 0.26)(d, 0.08)(e, 0.02))$$

$$(0.34, 0.30, 0.26, \quad 0.10 \quad)$$

$$0.08 \qquad 0.02$$

Un exemple de codage de Huffman

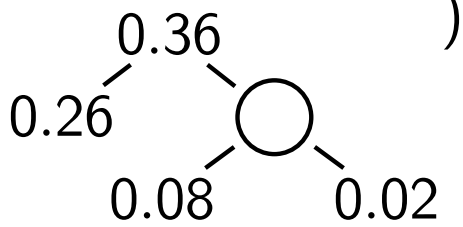
$$S = ((a, 0.34)(b, 0.30)(c, 0.26)(d, 0.08)(e, 0.02))$$

$$(0.34, 0.30, 0.26, \quad 0.10 \quad)$$

$$0.08 \qquad 0.02$$

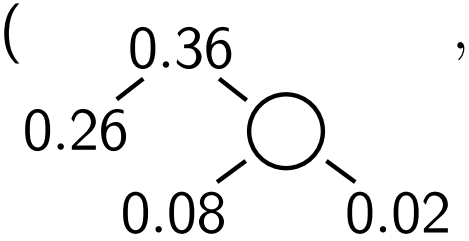
Un exemple de codage de Huffman

$$S = ((a, 0.34)(b, 0.30)(c, 0.26)(d, 0.08)(e, 0.02))$$

$(0.34, 0.30,$  $)$

Un exemple de codage de Huffman

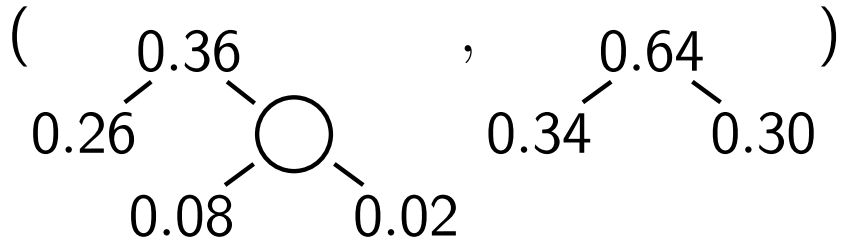
$$S = ((a, 0.34)(b, 0.30)(c, 0.26)(d, 0.08)(e, 0.02))$$

$$($$ 

$$, 0.34, 0.30)$$

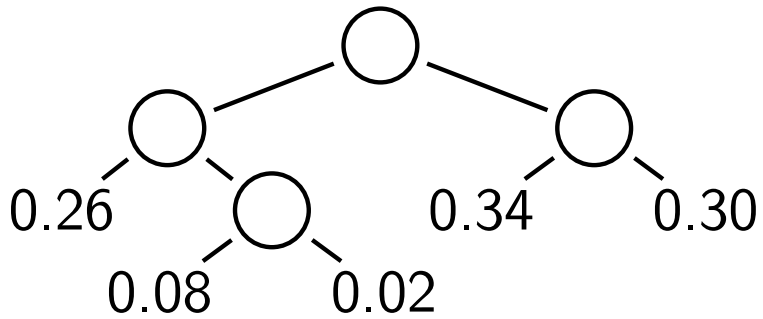
Un exemple de codage de Huffman

$$S = ((a, 0.34)(b, 0.30)(c, 0.26)(d, 0.08)(e, 0.02))$$



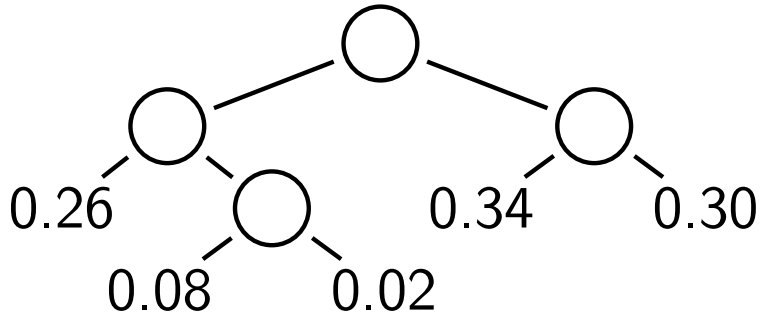
Un exemple de codage de Huffman

$$S = ((a, 0.34)(b, 0.30)(c, 0.26)(d, 0.08)(e, 0.02))$$



Un exemple de codage de Huffman

$$S = ((a, 0.34)(b, 0.30)(c, 0.26)(d, 0.08)(e, 0.02))$$



$$\begin{aligned} C(a) &= 10 \\ C(b) &= 11 \\ C(c) &= 00 \\ C(d) &= 010 \\ C(e) &= 011 \end{aligned}$$

Quelques remarques

- Un code binaire de Huffman a toujours une fonction de Kraft qui vaut 1 ; c'est également le cas de tout code optimal ($N \geq k$).

Quelques remarques

- Un code binaire de Huffman a toujours une fonction de Kraft qui vaut 1 ; c'est également le cas de tout code optimal ($N \geq k$).
- Si, au cours de l'algorithme de Huffman, on a à un certain moment des probabilités égales, il se peut que l'assortiment des longueurs (profil de l'arbre) ne soit pas unique.

Quelques remarques

- Un code binaire de Huffman a toujours une fonction de Kraft qui vaut 1 ; c'est également le cas de tout code optimal ($N \geq k$).
- Si, au cours de l'algorithme de Huffman, on a à un certain moment des probabilités égales, il se peut que l'assortiment des longueurs (profil de l'arbre) ne soit pas unique.
- On appelle parfois **codage de Huffman** d'un texte, le codage obtenu en prenant pour probabilités des différentes lettres leurs fréquences relatives d'apparition dans le texte.

Canal de transmission

- Un **canal de transmission** est "ce qui se passe" entre l'**émission** d'un message et sa **réception**

Canal de transmission

- Un **canal de transmission** est "ce qui se passe" entre l'**émission** d'un message et sa **réception**
- **Canal digital :** on décrit le message émis par une suite $X_1, X_2, \ldots$ de symboles, et le message reçu par une suite $Y_1, Y_2, \ldots$ de symboles du **même** alphabet (transmission sans codage).

Canal de transmission

- Un **canal de transmission** est "ce qui se passe" entre l'**émission** d'un message et sa **réception**
- **Canal digital :** on décrit le message émis par une suite $X_1, X_2, \ldots$ de symboles, et le message reçu par une suite $Y_1, Y_2, \ldots$ de symboles du **même** alphabet (transmission sans codage).
- **Bruit :** on s'attend à ce que l'on n'ait pas systématiquement $X_i = Y_i$

Canal de transmission

- Un **canal de transmission** est "ce qui se passe" entre l'**émission** d'un message et sa **réception**
- **Canal digital :** on décrit le message émis par une suite $X_1, X_2, \ldots$ de symboles, et le message reçu par une suite $Y_1, Y_2, \ldots$ de symboles du **même** alphabet (transmission sans codage).
- **Bruit :** on s'attend à ce que l'on n'ait pas systématiquement $X_i = Y_i$
- **Modélisation :** on fait l'hypothèse (très simplificatrice !) que le bruit s'applique indépendamment à chaque symbole et qu'il n'y a pas ni symboles perdus ni symboles surnuméraires ; le canal est alors modélisé par une "matrice de probabilités de transition"

$$p_{i,j} = \mathbb{P}(Y = s_j | X = s_i)$$

Information transmise

Si l'on *fixe* la loi de X, les caractéristiques du canal donnent la loi de (X, Y) et donc celle de Y.

Information transmise

Si l'on *fixe* la loi de X, les caractéristiques du canal donnent la loi de (X, Y) et donc celle de Y.
L'**information transmise** par le canal est représentée par l'**information mutuelle** $I(X, Y)$.

Information transmise

Si l'on *fixe* la loi de X, les caractéristiques du canal donnent la loi de (X, Y) et donc celle de Y.
L'**information transmise** par le canal est représentée par l'**information mutuelle** $I(X, Y)$.

- **Cas extrême :** bruit nul (transmission parfaite), $I(X, Y) = H(X) = H(Y)$; cela correspond à $Y = f(X)$, f inversible, et dans la pratique à $Y = X$.

Information transmise

Si l'on *fixe* la loi de X, les caractéristiques du canal donnent la loi de (X, Y) et donc celle de Y.
L'**information transmise** par le canal est représentée par l'**information mutuelle** $I(X, Y)$.

- **Cas extrême :** bruit nul (transmission parfaite), $I(X, Y) = H(X) = H(Y)$; cela correspond à $Y = f(X)$, f inversible, et dans la pratique à $Y = X$.
- **Cas intermédiaire** (cas pratique) : $0 < I(X, Y) < \min(H(X), H(Y))$

Information transmise

Si l'on *fixe* la loi de X, les caractéristiques du canal donnent la loi de (X, Y) et donc celle de Y.
L'**information transmise** par le canal est représentée par l'**information mutuelle** $I(X, Y)$.

- **Cas extrême :** bruit nul (transmission parfaite), $I(X, Y) = H(X) = H(Y)$; cela correspond à $Y = f(X)$, f inversible, et dans la pratique à $Y = X$.
- **Cas intermédiaire** (cas pratique) : $0 < I(X, Y) < \min(H(X), H(Y))$
- **Cas extrême :** aucune transmission d'information, $I(X, Y) = 0$: X et Y sont indépendants (on n'apprend rien sur le message émis, hormis sa longueur, en écoutant le message reçu).

Canal binaire symétrique

- **Canal binaire :** alphabet à 2 symboles, 0 et 1.

Canal binaire symétrique

- **Canal binaire :** alphabet à 2 symboles, 0 et 1.
- **Symétrique :** la probabilité d'erreur ϵ ne dépend pas du symbole émis

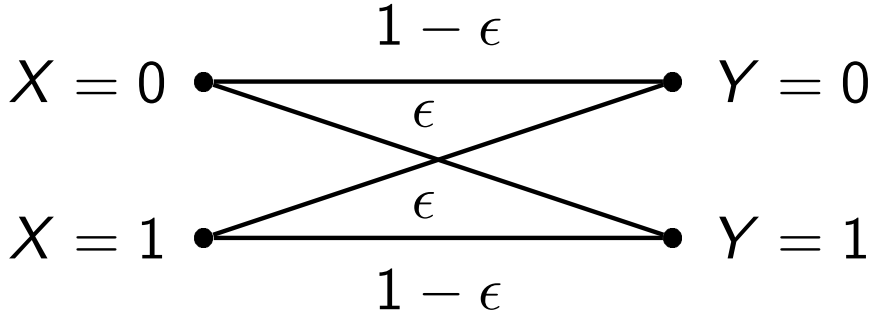
$$P = (p_{i,j})_{i,j=0,1} = \begin{pmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix}$$

Canal binaire symétrique

- **Canal binaire :** alphabet à 2 symboles, 0 et 1.
- **Symétrique :** la probabilité d'erreur ϵ ne dépend pas du symbole émis

$$P = (p_{i,j})_{i,j=0,1} = \begin{pmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix}$$

- Représentation schématique :



Capacité d'un canal

Définition : capacité d'un canal

Soit un canal fixé (défini par sa matrice de probabilités de transition). On appelle **capacité** du canal, le **maximum** de l'information mutuelle $I(X, Y)$, pris sur **toutes les probabilités possibles d'entrée**

$$C = \max_{(p_i)} I(X, Y)$$

Capacité d'un canal

Définition : capacité d'un canal

Soit un canal fixé (défini par sa matrice de probabilités de transition). On appelle **capacité** du canal, le **maximum** de l'information mutuelle $I(X, Y)$, pris sur **toutes les probabilités possibles d'entrée**

$$C = \max_{(p_i)} I(X, Y)$$

La capacité du canal représente la quantité maximale d'information que peut transmettre le canal, **pour peu** qu'on l'utilise avec les "bonnes" probabilités d'émission

Capacité d'un canal binaire symétrique

- Canal binaire symétrique avec taux d'erreur ϵ.

Capacité d'un canal binaire symétrique

- Canal binaire symétrique avec taux d'erreur ϵ.
- On doit calculer $I(X, Y) = H(Y) + H(X) - H(X, Y)$, en fonction de $p = \mathbb{P}(X = 1)$

Capacité d'un canal binaire symétrique

- Canal binaire symétrique avec taux d'erreur ϵ.
- On doit calculer $I(X, Y) = H(Y) + H(X) - H(X, Y)$, en fonction de $p = \mathbb{P}(X = 1)$
- **Astuce :** on pose $Z = 1$ si $X = Y$, $Z = 0$ sinon ; on a donc $\mathbb{P}(Z = 1) = 1 - \epsilon$, $\mathbb{P}(Z = 0) = \epsilon$.

Capacité d'un canal binaire symétrique

- Canal binaire symétrique avec taux d'erreur ϵ.
- On doit calculer $I(X, Y) = H(Y) + H(X) - H(X, Y)$, en fonction de $p = \mathbb{P}(X = 1)$
- **Astuce :** on pose $Z = 1$ si $X = Y$, $Z = 0$ sinon ; on a donc $\mathbb{P}(Z = 1) = 1 - \epsilon$, $\mathbb{P}(Z = 0) = \epsilon$.
- Le couple (X, Z) permet de retrouver (X, Y) (et réciproquement), donc $H(X, Y) = H(X, Z)$.

Capacité d'un canal binaire symétrique

- Canal binaire symétrique avec taux d'erreur ϵ.
- On doit calculer $I(X,Y) = H(Y) + H(X) - H(X,Y)$, en fonction de $p = \mathbb{P}(X = 1)$
- **Astuce :** on pose $Z = 1$ si $X = Y$, $Z = 0$ sinon ; on a donc $\mathbb{P}(Z = 1) = 1 - \epsilon$, $\mathbb{P}(Z = 0) = \epsilon$.
- Le couple (X, Z) permet de retrouver (X, Y) (et réciproquement), donc $H(X,Y) = H(X,Z)$.
- Par ailleurs (symétrie du canal), Z est indépendant de X, donc $H(X,Z) = H(X) + H(X)$.

Capacité d'un canal binaire symétrique

- Canal binaire symétrique avec taux d'erreur ϵ.
- On doit calculer $I(X,Y) = H(Y) + H(X) - H(X,Y)$, en fonction de $p = \mathbb{P}(X = 1)$
- **Astuce :** on pose $Z = 1$ si $X = Y$, $Z = 0$ sinon ; on a donc $\mathbb{P}(Z = 1) = 1 - \epsilon$, $\mathbb{P}(Z = 0) = \epsilon$.
- Le couple (X, Z) permet de retrouver (X, Y) (et réciproquement), donc $H(X,Y) = H(X,Z)$.
- Par ailleurs (symétrie du canal), Z est indépendant de X, donc $H(X,Z) = H(X) + H(X)$.
- Par conséquent, $I(X,Y) = H(Y) - H(Z)$; $H(Z)$ ne dépend pas de p, donc on cherche le p qui maximise $H(Y)$

Capacité d'un canal binaire symétrique

- Canal binaire symétrique avec taux d'erreur ϵ.
- On doit calculer $I(X,Y) = H(Y) + H(X) - H(X,Y)$, en fonction de $p = \mathbb{P}(X = 1)$
- **Astuce :** on pose $Z = 1$ si $X = Y$, $Z = 0$ sinon ; on a donc $\mathbb{P}(Z = 1) = 1 - \epsilon$, $\mathbb{P}(Z = 0) = \epsilon$.
- Le couple (X, Z) permet de retrouver (X, Y) (et réciproquement), donc $H(X,Y) = H(X,Z)$.
- Par ailleurs (symétrie du canal), Z est indépendant de X, donc $H(X,Z) = H(X) + H(X)$.
- Par conséquent, $I(X,Y) = H(Y) - H(Z)$; $H(Z)$ ne dépend pas de p, donc on cherche le p qui maximise $H(Y)$
- $\mathbb{P}(Y = 1) = p\epsilon + (1-p)(1-\epsilon)$, vaut $1/2$ pour $p = 1/2$, et alors

$$C = I(X,Y)_{p=1/2} = 1 - H(Z) = 1 + \epsilon \log \epsilon + (1-\epsilon)\log(1-\epsilon)$$