Fiche de Théorie de l'Information

notes par Claire Davin, mise en page May Cattant

Contents

Part I
Mesure de l'information

1 Pour des évènements

Information apportée par un évènement E (incertitude) $h(E) = \log_2 \frac{1}{P(E)}$ [bits]

Incertitude conditionnelle $h(E|F) = \log_2 \frac{1}{P(E|F)}$ [bits]

Incertitude liée à la réalisation simultanée de deux évènements $h(E \cap F) = \log_2 \frac{1}{P(E \cap F)} = \log_2 \frac{1}{P(E|F)P(F)} = h(E|F) + h(F)$

Information apportée par F sur E $I_{F \to E} = h(E) + h(E|F) = \log_2 \frac{P(E|F)}{P(E)}$

- $I_{F \to E} > 0 \leftrightarrow P(E|F) > P(E)$ lorsque F est réalisée, cela élimine des possibilités défavorables à la réalisation de E
- $I_{F \to E} < 0 \leftrightarrow P(E|F) < P(E)$ lorsque F est réalisée, cela élimine des possibilités favorables à la réalisation de E

Propriétés

- $h(E \cap F) = h(E) + h(F) - I(E, F)$
- si E et F indépendants, $I_{F \to E} = 0$
- $I_{E \to F} = I_{F \to E}$ information mutuelle notée $I(E, F)$

2 Pour des variables aléatoires

Notations

- X v.a. discrète à valeurs dans $x_1, ..., x_n$ avec $p_i = P(X = x_i)$
- Y v.a. discrète à valeurs dans $y_1, ..., y_n$ avec $q_i = P(Y = y_i)$
- $p_{ij} = P(X = x_i \cap Y = y_i)$

Entropie de X nombre d'éléments binaires moyen nécessaire pour encadrer les valeurs prises par X
$H(X) = -\sum_{i=1}^{n} p_i \log p_i$ = incertitude moyenne

Entropie du couple $H(X, Y) = -\sum_{i=1}^{n} \sum_{j=1}^{n} p_{ij} \log p_{ij}$

Entropie conditionnelle de X sachant $Y = y_j$ $H(X|Y = y_j) = -\sum_{i=1}^{n} P(X = x_i|Y = y_j) \log(P(X = x_i|Y = y_j))$

Entropie conditionnelle de X sachant Y $H(X|Y) = +\sum_{j=1}^{n} q_j H(X, Y = y_j)$

Information moyenne apportée par Y sur X est symétrique, à choisir intelligemment

$$\begin{aligned} I_{Y\to X} &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \end{aligned}$$

Propriétés

- $H(X) \leq \log n$
- $I_{Y\to X} \geq 0 \Leftrightarrow H(X|Y) \leq H(X)$ càd le conditionnement diminue l'incertitude moyenne
- X et Y indépendantes $\Rightarrow I(X,Y) = 0$

Part II
Sources discrètes et algorithmes de compression

1 Introduction

On dispose d'une réalisation d'une variable U dans le temps

- U est stationnaire $\Leftrightarrow$ les u_k ont même loi de probabilité
- U a une mémoire d'ordre m $\Leftrightarrow P(\{u_n = i_n\} | \cap_{j=0}^{n-1} u_j = i_j) = P(\{u_n = i_n\} | \cap_{j=n-m}^{n-1} u_j = i_j)$
  - $m = 0 \Rightarrow$ source sans mémoire (indépendance des u_k
  - $m = 1 \Rightarrow$ chaîne de Markov

2 Codage de source

Hiérarchie des codes 1>2>3>4

1. codes quelconques $x \to c(x)$
2. codes non singuliers (injectifs) $x_1 \neq x_2 \Rightarrow c(x_1) \neq c(x_2)$
3. codes uniquement décodables/déchiffrables
4. codes préfixes : code qui ne possède aucun mot ayant pour préfixe un autre mot

Théorème de Kraft un code C peut être transformé en code préfixe équivalent si et seulement si $\sum_{c\in C} b^{-n(C)} \leq 1$ où b est la taille de l'alphabet utilisé et $n(C)$ longueur des mots

Théorème de Mac-Millan un code uniquement déchiffrable vérifie $\sum b^{-n(C)} \leq 1$

Théorème du codage de source Pour une source stationnaire U

$$\underbrace{\frac{H(L)}{\log b} + \frac{1}{L} > \overline{n}}_{\text{il existe un code préfixe pour coder une extension d'ordre L de la source tel que}} = \underbrace{\frac{\overline{n_L}}{L} \geq \frac{H_L}{\log b}}_{\text{Tout codage par un code uniquement déchiffrable vérifie}}$$

- $\overline{n_L}$ nombre moyen de symboles code correspondant à un mot source de longueur L
- $\overline{n}$ nombre moyen de symboles code nécessaires à la représentation un symbole source
- $H(L) = \frac{H(u_1,\ldots,u_n)}{L}$

Part III
Canaux discrets

1 Capacité d'un canal

Information moyenne apportée par la sortie sur l'entrée $I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$

Capacité $C = maxI(X,Y)$

Matrice de transition $T = [t_{ij}]_{1 \leq i,j \leq n}$ où $t_{ij} = P(Y = y_i | X = x_i)$

Différents types de canaux

- canal discret : alphabets de sortie et d'entrée ne comprennent qu'un nombre fini de symboles
- canal sans mémoire : la sortie y_k ne dépend que de l'entrée x_k
- canal symétrique : il existe une partition de l'alphabet de sortie telle que les sous-matrices de transition correspondantes ont leurs lignes (et colonnes) identiques à une permutation près

Calcul de la capacité Si le canal est symétrique, la capacité est atteinte pour une loi uniforme sur l'entrée

2 Second théorème de Shannon

$\forall \epsilon > 0$, il existe un codage canal tel que la probabilité d'erreur de transimission $P_e < \epsilon$ à condition que $H' = D_S.H_\infty(S) < D_C.C = C'$

- H' débit d'entropie de la source
- D_S débit de la source [lettre/s]
- $H_\infty(S)$ entropie de la source par symbole [bits]
- D_c débit du canal [symbole transmis/s]
- C capacité du canal [bits/symbole transmis]
- C' capacité du canal par unité de temps

Part IV
Codes correcteurs d'erreur

1 Généralités

Mot de code $\underbrace{a_1 a_2 .... a_m}_{\text{m bits d'information}} \underbrace{a_{m+1} ... a_{m+k}}_{\text{k bits de contrôle}}$

Décodage optimal au sens du maximum de vraissemblance = "décodage à distance minimum". On cherche le mot $\hat{c}$ tel que $P(Y|C)$ est maximum

2 Détection d'un code

Matrice génératrice du code C $G = \begin{Bmatrix} Id_m \\ P \end{Bmatrix}$ m colonnes, n lignes

Matrice de contrôle on a $H = \left\{ \begin{array}{c} P^T \\ Id_{n-m} \end{array} \right\}$ Y est un mot de code $\Leftrightarrow H^T Y = 0$

Pouvoir correcteur syndrôme $s(Y) = H^T Y$

Codage de Huffman Algorithme de compressionde code (voir TD3)

- On classe par ordre de probabilité croissante
- on relie les symboles de probabilités les plus faibles (arbre)
- on obtient le nouveau code pour chaque mot source en partant de la racine