

Chapitre 1

L'information et le codage

1. Historique et introduction

La théorie de l'information est née dans le contexte de la théorie statistique des communications. Ses méthodes, essentiellement mathématiques, ont permis de rendre compte et d'expliquer l'évaluation des performances des systèmes de communications, en raisonnant au niveau le plus élémentaire, abstraction faite des moyens ou supports physiques.

- 1928, Hartley : 1^o tentative de définition scientifique d'une « quantité d'information »
- 1948, C. Shannon : introduit le nouveau concept de « quantité d'information » de façon mathématique, en déduisant les principales conséquences : réel début de la « théorie de l'information ». Référence : Claude E. Shannon, A Mathematical Theory of Communication, Bell System Technical Journal, vol. 27, pp.379-423 and 623-656, July and October 1948, <http://pespmc1.vub.ac.be/books/Shannon-TheoryComm.pdf>

- La théorie de l'information se préoccupe des systèmes de communication et de leur efficacité.
- Il s'agit d'une théorie mathématique qui décrit les aspects les plus fondamentaux des systèmes de communications.
- Cette théorie s'intéresse à la construction et à l'étude de modèles mathématiques à l'aide essentiellement de la théorie des probabilités.
- Elle consiste en l'élaboration et l'étude de modèles pour la source et le canal qui utilisent différents outils comme les probabilités, les automates finis, etc.

La théorie des communications s'intéresse aux moyens de transmission d'une information depuis une source jusqu'à un récepteur à travers un canal.

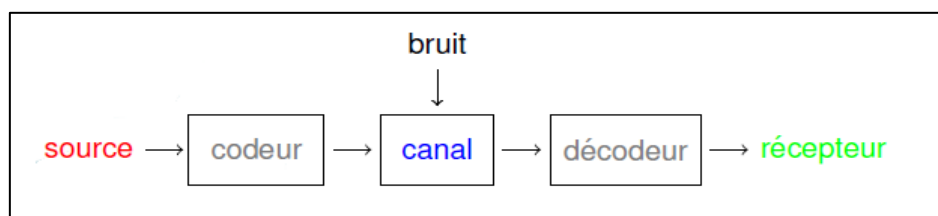


Figure 1.1. Schéma d'un système de communication

1.1. La source : Générateur de message, elle fournit un signal, la nature de la source peut être très variée : un signal électromagnétique, une séquence de symbole binaire ou une voix.

Nous distinguons deux types de source : analogique et discrète.

- Un microphone est une source d'informations analogique produisant un signal analogique en amplitude et en temps.
- Une source d'information discrète produit des symboles (ou des lettres comme un clavier). La sortie est une chaîne (ou une séquence) de symboles.

Exemple

Un dé par exemple, choisit le symbole à sortir dans l'ensemble de $\{1, \dots, 6\}$. Une séquence possible peut être $\{1,5,4,1,2,4,1,1\}$.

La question essentielle est la suivante : quel est le montant de l'information émise par cette source ? Et ensuite, quelle unité de mesure choisit-on pour l'information ?

1.2. Le canal : Le canal peut être une ligne téléphonique, une liaison radio ou encore un support magnétique ou optique : bande magnétique ou disque compact; La transmission peut se faire dans l'espace ou dans le temps ;

1.3. Bruit : Le canal sera généralement perturbé par un bruit qui dépendra de l'environnement et de la nature du canal : perturbations électriques, rayures perturbateur du canal...

1.4. Le codeur : Le codeur représente l'ensemble des opérations effectuées sur la sortie de la source avant la transmission. Ces opérations peuvent être, par exemple, la modulation, la compression ou encore l'ajout d'une redondance pour combattre les effets du bruit. Elles ont pour but de rendre la sortie de la source compatible avec le canal.

1.5. Le décodeur : le décodeur devra être capable, à partir de la sortie du canal de restituer de façon acceptable l'information fournie par la source.

2. Mesure de l'information

2.1. Notions de quantité d'information

2.1.1. Notion d'information

- L'information, comme l'énergie d'ailleurs, est une notion fondamentale, qui fait partie de notre vie quotidienne.
- L'information peut être transportée (transmission), stockée (mémorisation), transformée (traitement).
- L'information permet de diminuer l'incertitude.

Exemples

1. Considérons par exemple une source qui peut produire trois symboles a, b et c. Quand le destinataire attend un symbole, il est dans l'incertitude quant au symbole que la source va engendrer. Lorsque le symbole apparaît et qu'il arrive au destinataire, cette incertitude diminue.

2. Il neige en janvier contient beaucoup moins d'information que Il neige en août pour peu que l'on soit dans l'hémisphère nord.

3. On recherche une lettre dans une boîte. Si on précise que la lettre se trouve dans une enveloppe **bleue**, on fournit une information qui diminuera le temps de recherche du fait que le nombre de lettres dans des enveloppes bleues est plus restreint.

Si on ajoute l'information que la lettre dans une grande enveloppe, on pourra abréger d'autant plus le temps de la recherche.

Le but de la théorie de l'information est de mesurer cette incertitude avant réception.

2.2. Quantité de l'information

Nous essayons d'inventer une relation mathématique pour quantifier l'information. Supposons qu'une source d'information émet un des messages possibles $m_1, m_2, \dots, m_q$. Avec les probabilités de réalisation $p_1, p_2, \dots, p_q$ ($p_1 + p_2 + \dots + p_q = 1$). Par intuition, on sait que l'information portée par $k^{\text{ième}}$ message ($I(m_k)$) doit être inversement proportionnelle à p_k . L'autre critère par intuition est que quand p_k tend vers 1, la quantité de l'information tend vers zéro. On sait aussi que $I(m_k)$ est toujours positif. C'est-à-dire :

$$I(m_k) > I(m_j) \text{ si } p_k < p_j$$

$$I(m_k) \rightarrow 0 \text{ quand } p_k \rightarrow 1$$

$$I(m_k) \rightarrow 0 \text{ quand } 0 < p_q < 1$$

Avant de trouver une fonction satisfaisant toutes ces contraintes, nous allons imposer une autre. Si nous avons deux messages indépendants, l'information contenue dans les deux est la somme de l'information de chacun. Cette nouvelle contrainte est parfaitement justifiée par l'intuition. Une fonction logarithmique satisfait toutes les requis ci-dessus. Ainsi nous définissons le montant de l'information contenue dans un message qui se produit avec une possibilité p_k comme suit :

$$I(m_k) = \log\left(\frac{1}{p_k}\right)$$

L'information $I(m_k)$ apportée par la réalisation de m_k d'autant plus grande qu'elle est improbable. Elle peut s'exprimer ainsi :

$$I(A) = f\left(\frac{1}{P(m_k)}\right)$$

La fonction $f(.)$ vérifie les contraintes suivantes :

- $f(.)$ est croissante
- Information apportée par 1 événement sûr est nulle : $\lim_{p \rightarrow 1} [f(p)] = 0$
- Information apportée par 2 événements indépendants : $f(p_1 \cdot p_2) = f(p_1) + f(p_2)$

(La quantité d'information de deux symboles (caractères) successifs est la somme de leurs quantités d'information)

Ceci nous conduit à utiliser la fonction logarithmique pour $f(.)$

On aboutit résultat escompté, soit

$$f(p) = -\log_b p$$

Définition. Soit (Ω, A, P) un espace probabilisé et A un événement de probabilité $P(A)$ non-nulle. On associe à la réalisation de A la quantité d'information propre :

$$I(A) = -\log P(A) = \log\left(\frac{1}{P(A)}\right) = \log\left(\frac{1}{P_A}\right)$$

L'information ainsi définie est une grandeur toujours positive. Si la source émet le symbole A avec la probabilité $p=1$ (événement certain), on a une quantité d'information nulle $I(A) = 0$ associée à cet événement certain.

2.3. Unités

L'unité dépend de la base choisie pour le logarithme.

$\log_2$: Shannon, bit (binary unit)

$\log_e$: logon, nat (natural unit)

$\log_{10}$: Hartley, decit (decimal unit)

Exemples :

- Une source met en sortie un des 5 symboles possibles dans chaque intervalle de symbole.

La probabilité de réalisation de chaque symbole est donnée ci-dessous :

$$P_1 = \frac{1}{2}, P_2 = \frac{1}{4}, P_3 = \frac{1}{8}, P_4 = \frac{1}{16}, P_5 = 1/16$$

Calculer le contenu de l'information concernant chaque message (symbole) ?

Solution :

$$I(m_1) = \log((1 / (1/2))) = 1 \text{ bit}$$

$$I(m_2) = \log((1 / (1/4))) = 2 \text{ bits}$$

$$I(m_3) = \log((1 / (1/8))) = 3 \text{ bits}$$

$$I(m_4) = \log((1 / (1/16))) = 4 \text{ bits}$$

$$I(m_5) = \log((1 / (1/16))) = 4 \text{ bits}$$

Vocabulaire : $I(\cdot)$ est désigné par incertitude ou encore quantité d'information.

3. Quantité d'information conditionnelle

La définition de la quantité d'information propre s'applique à la réalisation conjointe de A et B. En remarquant que $P(A \cap B) = P(A) P(B|A)$, on obtient :

$$I(A \cap B) = -\log P(A \cap B) = -\log P(A) - \log P(B|A)$$

On note que $-\log P(B|A)$ correspond à la quantité d'information propre de B que ne fournit pas l'observation de A.

Définition : On appelle information conditionnelle de B sachant A la quantité :

$$I(B|A) = -\log P(B|A);$$

Soit, en d'autres termes :

$$I(B|A) = I(A \cap B) - I(A)$$

4. Information propre et mutuelle (Quantité d'information mutuelle)

La définition de l'information conditionnelle amène directement une autre définition, celle de l'information mutuelle, qui correspond à la part d'incertitude commune à deux événements.

Définition 3. On appelle information mutuelle entre A et B la quantité :

$$I(A, B) = I(A) - I(A|B) = I(B) - I(B|A)$$

5. Entropie

On transmet habituellement sur un canal de communication des séquences de symboles de grande longueur émises par la source d'information. On s'intéresse donc plus au contenu moyen d'information produit par une source qu'au contenu informatif d'un seul symbole.

Soit une source S d'information sans mémoire sélectionnant aléatoirement un symbole parmi les n éléments d'un alphabet $\{s_1, \dots, s_n\}$. Soit p_i la probabilité d'apparition de s_i . La quantité d'information moyenne associée à l'apparition de chaque symbole possible est donnée par :

$$H(s) = E[I(s)] = \sum_{i=1}^n p_i I(s) = - \sum_{i=1}^n p_i \log p_i = \sum_{i=1}^n p_i \log \left(\frac{1}{p_i} \right)$$

Bits/symbole ou Shannon/symboles

Exemple (d'une variable aléatoire binaire)

Source binaire avec deux symboles « 0 » et « 1 »

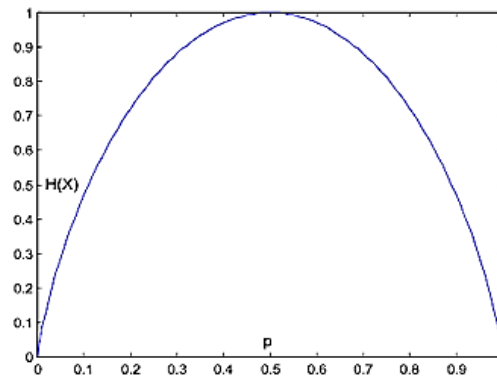
$$0 \rightarrow P$$

$$1 \rightarrow 1 - P$$

L'entropie de cette source :

$$H(S) = \sum_{i=1}^2 p_i \log \left(\frac{1}{p_i} \right) = P \log \left(\frac{1}{P} \right) + (1 - P) \log \left(\frac{1}{1 - P} \right)$$

$$H(S) = -P \log p - (1 - P) \log (1 - P)$$



La fonction $H(S)$ est symétrique par rapport à $p=0.5$ et est maximale pour $p=0.5$

Remarque

L'entropie d'une source est maximale lorsque tous les symboles sont équiprobables donc :

$$P(x_i) = \frac{1}{n}$$

$$H(S) = \sum_{i=1}^n \frac{1}{n} \log(n) = \log(n)$$

$$0 \leq H(S) \leq \log(n)$$

5.1. Entropie conjointe

Si $p(x_i, y_j)$ est la densité de probabilité jointe entre deux caractères alors la quantité d'information jointe est :

$$I(x_i, y_j) = P(x_i, y_j) \log \frac{1}{P(x_i, y_j)}$$

Soient X et Y des variables aléatoires à valeurs dans $\{x_1, \dots, x_n\}$ et $\{y_1, \dots, y_m\}$, respectivement.

L'entropie conjointe de X et Y est donnée :

$$H(X, Y) \triangleq - \sum_{i=1}^n \sum_{j=1}^m P(X = x_i, Y = y_j) \log P(X = x_i, Y = y_j)$$

L'entropie jointe des deux sources est alors la quantité d'information moyenne jointe entre deux caractères de la source :

Remarque

L'entropie conjointe est une grandeur symétrique c'est-à-dire : $H(X, Y) = H(Y, X)$

- **Cas où les sources sont indépendantes :**

$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$$

$$H(X, Y) = H(X) + H(Y)$$

- **Cas ou les sources sont dépendantes :**

Quantité d'information mutuelle

$$H(X, Y) = \sum_{i=1}^n \sum_{j=1}^m P(X = x_i, Y = y_j) \log \left(\frac{1}{P(X = x_i, Y = y_j)} \right) = \sum_{ij} P(x_i, y_j) \log \left(\frac{1}{P(x_i, y_j)} \right)$$

La quantité d'information mutuelle :

$$I(x, y) = \sum_{ij} P(x_i, y_j) \log \left(\frac{P(x_i, y_j)}{P(x_i) P(y_j)} \right)$$

$$I(x, y) = \sum_{i=1}^n \sum_{j=1}^m P(X = x_i, Y = y_j) \log \left(\frac{P(X = x_i, Y = y_j)}{P(X = x_i) P(Y = y_j)} \right)$$

L'information mutuelle représente l'incertitude moyenne sur X diminuée de celle qui subsiste lorsque Y est connue.

Donc :

$$H(X, Y) = H(X) + H(Y) - I(x, y)$$

C'est l'entropie conjointe de deux sources dépendantes

5.2. Entropie conditionnelle

Définition 6. Soient X et Y des variables aléatoires à valeurs dans $\{x_1, \dots, x_n\}$ et $\{y_1, \dots, y_m\}$.

L'entropie conditionnelle de X sachant que $Y = y_j$ est donnée :

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

Incertainité moyenne sur le résultat de X sachant que celui de Y est y_j .

Définition 7. L'entropie conditionnelle moyenne de X sachant Y est donnée par :

$$H(X|Y) \triangleq \sum_{j=1}^m P(Y = y_j) H(X|Y = y_j)$$

Remarque

L'entropie conditionnelle permet d'obtenir d'autres formulations de ces qualités en utilisant la loi de BAYES :

$$P(x, y) = P(X|Y) P(Y) = P(Y|X) P(X)$$

$$I(X, Y) = \sum_{ij} P(x_i, y_j) \log \left(\frac{P(x_i, y_j)}{P(x_i) P(y_j)} \right)$$

$$I(x, y) = -H(x|y) + H(x)$$

$$I(x, y) = -H(y|x) + H(y)$$