Introduction aux systèmes de communication

1 Introduction

La théorie des communications s'intéresse aux moyens de transmettre une information depuis la source jusqu'à un utilisateur à travers un canal.
La nature de la source peut être très variée. Il peut s'agir par exemple d'une voix, d'un signal électromagnétique ou d'une séquence de symboles binaires.
Le canal peut être une ligne téléphonique, une liaison radio, un support magnétique ou optique. La transmission peut se faire dans l'espace ou dans le temps. 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, le brouillage, l'ajout de redondance pour combattre les effets du bruit, ou encore l'adaptation à des contraintes de spectre. Elles ont pour but de rendre la sortie de la source compatible avec le canal. Enfin le décodeur doit être capable, à partir de la sortie du canal, de restituer de façon acceptable l'information fournie par la source.

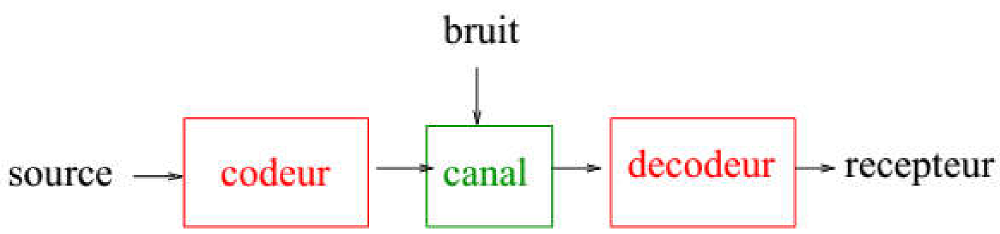

Figure 1 – Un système de communication

La théorie de l'information a été créée par C. E. Shannon dans les années 40. Il s'agit d'une théorie mathématique qui décrit les aspects les plus fondamentaux des systèmes de communication. Elle consiste en l'élaboration et l'étude de modèles pour la source et le canal qui utilisent différents outils comme les probabilités et les automates finis.

– Le but du codeur de source est de représenter la sortie de la source en une séquence binaire, et cela de façon la plus économique possible.
– Le but du codeur de canal et de son décodeur est de reproduire le plus fidèlement possible cette séquence binaire malgré le passage à travers le canal bruité.

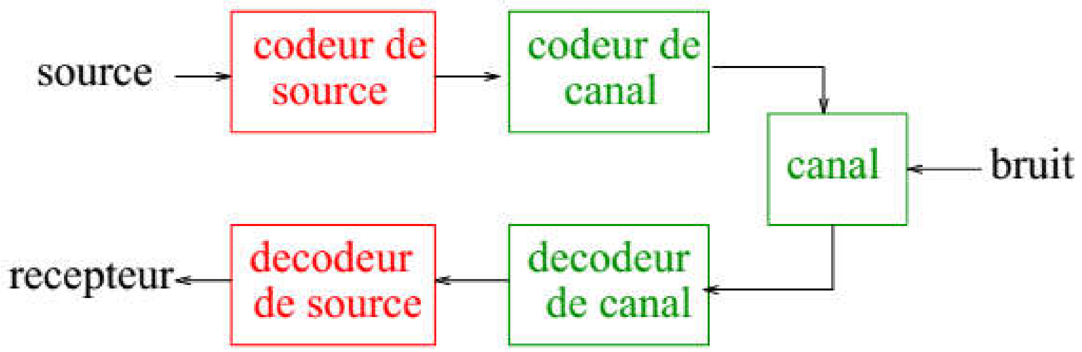

Figure 2 – Codeur de source et codeur de canal

Cette séparation entre codage de source et codage de canal n'implique en pratique aucune limitation sur les performances du système complet.

2 Sources et codage de source

Formellement, une source d'information *S* a pour sortie une suite d'échantillons ou de symboles *A*= (*a*1, . . . , *an*). Chaque symbole *xi* à une probabilité p(*ai*) d'apparaitre. Le codeur de source code chaque entrée a*i* par un mot de code *ci* sous forme binaire. Le décodeur de source récupère chaque forme binaire au niveau du destinataire et fournit un bloc *y* l'image de la source reconstruite pour le destinataire.

La sortie d'une telle source est une séquence de lettres tirées dans un alphabet fini A = {a1, . . . , an}. Chaque lettre de la séquence est choisie aléatoirement d'après une loi de probabilité p indépendante du temps. Pour toute lettre a, p(a) est la probabilité pour que cette lettre soit choisie. Il s'agit d'un réel compris entre 0 et 1. On a $\sum a \in A \; P(a) = 1$. La donnée de p(a1), . . . , p(an) définit la probabilité discrète p sur A.
Il peut sembler étonnant de modéliser une source d'information à l'aide d'une variable aléatoire. Nous allons donner un exemple qui permet de se convaincre de l'utilité de tels modèles.
Exemple : Soit une source d'information qui fournit comme information l'une des quatre lettres a1, a2, a3, a4. Supposons que le codage de source transforme cette information discrète en symboles binaires. Nous donnons deux exemples de codage différents.

Codage 1	Codage 2
$a_1 \to 00$	$a_1 \to 0$
$a_2 \to 01$	$a_2 \to 10$
$a_3 \to 10$	$a_3 \to 110$
$a_4 \to 11$	$a_4 \to 111$

Si les quatre lettres sont équiprobables, la première méthode de codage est meilleure. Elle nécessite en effet deux symboles par lettre en moyenne tandis que la deuxième méthode nécessite $\frac{1}{4} + 2 * \frac{1}{4} + 3 * \frac{1}{4} + 3 * \frac{1}{4} = 2.25\ symboles\ par\ lettres.$

En revanche, si l'on a une source dont la distribution de probabilités est
p(a1) = 1/2, p(a2) = 1/4, p(a3) = p(a4) = 1/8,
la longueur moyenne d'un symbole codé par la première méthode est toujours 2 tandis que celle d'un symbole codé par la deuxième méthode est

$$\frac{1}{2} + 2 * \frac{1}{4} + 3 * \frac{1}{4} + 3 * \frac{1}{8} = 1.75\ .$$

Le deuxième codage réussit donc à coder quatre symboles avec moins de deux bits.
Pour coder correctement une source, il est donc important de connaître son comportement statistique.

3. Information propre et entropie

3.1. Information propre

Mesurer la quantité d'information que contient un message a toujours été est un problème difficile. L'information n'est pas une entité physique mais un concept abstrait et difficile à quantifier surtout quant un facteur humain est inclus.
Pour surmonter ce problème, Shannon a donné l'idée de définir le contenu en information d'un événement comme une fonction *h(x)* qui dépend seulement de sa probabilité. Il a mesuré l'information apportée par un événement E par le logarithme de l'inverse de sa probabilité de réalisation. Donc, soit deux v.a.: *X* avec une distribution $p_X(x)$ (ou seulement *p(x))* et *Y* avec une distribution $p_Y(y)$. La v.a. *X* prend ces valeurs dans un alphabet X et *Y* prend ces valeurs dans Y.
Pour choisir la mesure de l'information, on ajoute les axiomes suivants:
-cette fonction *h(x)* doit être décroissante. Parce que le plus un événement est probable moins il.
- un événement sûr ne porte aucune information. Si *p(x)=1, h(x)=0*.
- Claude Shannon a mesuré l'information apportée par un événement E par le logarithme de l'inverse de sa probabilité de réalisation
- pour deux événements indépendants, l'information totale est la somme de l'information de chaque événement.
La seule fonction qui satisfait ces axiomes est la fonction logarithmique. Donc:

$$h(x) = \log \frac{1}{p(x)} = -\log(p(x))$$

Définition 1

L'information propre est définie par :

$$I(E) = \log\left(\frac{1}{p(E)}\right) = -\log\left(p(E)\right)$$

Cette quantité s'appelle information propre (self-information) et elle correspond à la surprise ressentie lors de la réalisation d'événement peu probable.

L'unité de l'entropie dépend de la base du logarithme utilisée dans son calcul. Dans le cas numérique la base binaire impose un logarithme à base de 2 et l'unité de l'information propre devient ainsi le bit.

Exemple

Considérons un jeu de 32 cartes. Si on extrait une carte au hasard, quelle est l'incertitude liée à l'événement E= la carte est le roi de cœur? En supposant, ce qui ne semble pas déraisonnable, que chacune des 32 cartes a la même probabilité 1/32 d'être extraite,

la réponse est: $I(E) = \log\left(\frac{1}{p(E)}\right) = log32 = 5bits$

Remarque *I*(*E*) vaut zéro si *p*(*E*) = 1.

- L'information propre est une grandeur additive : si les évènements x et y sont statistiquement indépendants alors l'information totale qu'ils peuvent fournir est la somme des informations propres :
I(x,y)=f(1/p(x, y)) = f(1/p(x).1/p(y)) = f(1/p(x)) + f(1/p(y))=I(x)+I(y)

Donc pour deux v.a. indépendants *x* et *y*, l'information totale obtenue est :

$$h(x;y) = \log \frac{1}{p(x;y)} = \log \frac{1}{p(x)p(y)} = \log \frac{1}{p(x)} + \log \frac{1}{p(y)}$$

Donc elle satisfit la relation: h(x,y)=h(x)+h(y)

Exemple
_ Soit une source dont l'alphabet de sortie *{a0,…,a15}* avec *P(ak)=1/16*.
_ L'information propre de l'une de ces sorties *ak* est égale à : *I(ak)*=log2(16)=4 bits
_ Information : choisir k dans {0,…,15} => besoin de 4 bits

_ Attention : ce résultat = vrai car équiprobabilité

3.2. Entropie

La valeur moyenne de l'information propre calculée sur l'ensemble de symboles de S a une grande importance. Elle indique l'espérance de la variable aléatoire *S*. Elle est appelée entropie de la source et est notée *H*(*S*) :

Définition 2

L'entropie d'une source est définie par

$H(S) = -\sum_{i=0}^{N} P(Xi\,)\log\,(P(Xi)$

L'unité de l'entropie dépend de la base du logarithme utilisée dans son calcul. Dans le cas numérique la base binaire impose un logarithme à base de 2 et l'unité de l'entropie devient ainsi le bit.

Caractéristique de l'entropie

- Comportement probabiliste moyen de la source:
    - La source est une variable aléatoire X qui réalise les événements (émet les symboles) xi.
    - Elle est discrète, finie et ... stationnaire :

    pi= P(X=xi) pour i de 1 à n et Σ pi = 1
- La quantité d'information moyenne pour chaque xi est la moyenne de L'information de chaque événement X = xi :

$$H(X) = E[I(X)] = \sum_{i=1}^{n} p_i\, I(x_i) = \sum_{i=1}^{n} p_i \log(1/p_i)$$

- H(X) est **l'entropie de la source X** (entropie moyenne par symbole)

Exemple

Soit une source S= {x1, x2, x3, x4} Avec p(x1)=1/2; p(x2)=1/4 et p(x3)=p(x4)=1/8. L'entropie de S est :

$$H(S) = -\frac{1}{2log2\left(\frac{1}{2}\right)} - \frac{1}{4log2\left(\frac{1}{4}\right)} - 2 * \frac{1}{8log2\left(\frac{1}{8}\right)} = 7/4bits$$

Dans le cadre précis des code non ambigus, l'entropie s'avère être la limite fondamentale à la compression de données, c'est-à-dire qu'elle est fortement liée à la longueur moyenne de code minimale possible.

3.3. Incertitude et information

Qualitativement, fournir une information consiste donc à lever une incertitude sur l'issue d'une expérience aléatoire. La notion d'information est déjà inhérente à celle de probabilité conditionnelle. Considérons les événements {A = a} et {B = b}. La probabilité
p(a|b) peut être interprétée comme la modification apportée à la probabilité
p(a) de l'événement {A = a} lorsque l'on reçoit l'information que l'événement {B = b} s'est réalisé. Ainsi

– si $p(a|b) \le p(a)$, l'incertitude sur a augmente,

– si $p(a|b) \ge p(a)$, l'incertitude sur a diminue.

Pour mesurer la variation de l'incertitude, il faut choisir une fonction dé-croissante de la probabilité. On choisit également une fonction continue. Le logarithme permet d'exprimer commodément les variations d'incertitude. On notera I(a) l'incertitude sur a, encore appelée information propre de a :

$$I(a) = -log2p(a)$$

_ Information propre : *I(x) = log(1/p(x)) = -log(p(x))*

_ Information conjointe : *I(x,y) = log (1/p(x, y)) = - log (p(x,y))*

_ Information conditionnelle : *I(x | y) = log (1/p(x | y)) = -log(p(x | y))*

_ La règle de Bayes : *P(x,y) = P(x|y).P(y) = P(y|x).P(x)* donne

I(x,y) = I(y) + I(x|y) = I(x) + I(y|x)

_ Information mutuelle :

I(x;y) =log(p(x | y)/p(x))=log(p(x,y)/(p(x)p(y))) = log(p(y | x)/p(y)) = I(y;x)

I(x;y) = I(x) - I(x|y)

_ *Si I(x;y) > 0 =>* si un des événements se réalise, alors la probabilité d'occurrence de l'autre augmente. (si *I(x;y) < 0 …)*

_ *Si I(x;y) = 0 =>* les deux événements sont statistiquement indépendants.

Ainsi l'information «b est réalisé» diminue l'incertitude sur a de la quantité :

$$I(a) - I(a\backslash b) = log2p(a\backslash b)/p(a)$$

Cette dernière quantité est appelée information mutuelle de a et b.

Exemple : Soit A = {a0, . . . , a15} un alphabet de 16 lettres équiprobables. L'information propre d'une lettre a quelconque est I(a) = −log2(1/16) = 4. Dans ce cas particulier, l'information va consister à choisir un entier i dans
{0,1, . . . ,15} et pour représenter cette information il faut disposer de 4 bits.

Il faut prendre garde au fait que ceci n’est vrai que parce que les lettres sont équiprobables. En effet, si ce n’est pas le cas, l’information propre d’une lettre sera généralement différente de 4 et l’information propre moyenne peut même être strictement inférieure à 4.

3.4. Information mutuelle et entropie jointe

La formule de l’entropie d’une variable aléatoire simple peut être étendue à deux variables aléatoires formulant l’entropie jointe, l’entropie conditionnelle et l’information mutuelle.
Définition

L’entropie jointe est définie par

$$H(x,y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log\left(p(x,y)\right)$$

Définition L’entropie conditionnelle est définie par

$$H(x \backslash y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log\left(p(x \backslash y)\right)$$

L’information mutuelle quantifie l’information qu’une variable aléatoire partage avec une autre variable aléatoire et elle se mesure par la formule :

Définition 2.5 L’information mutuelle est définie par

$$\boldsymbol{I(x,y)} = \sum_{x \in X} \sum_{y \in Y} \boldsymbol{p(x,y)}\ \boldsymbol{log}\left(\frac{\boldsymbol{p(x,y)}}{\boldsymbol{p(x)p(y)}}\right) = \boldsymbol{H(y)} - \boldsymbol{H(Y \backslash X)}$$

Propriétés

- L’entropie H est positive ou nulle : $H(p1,\ldots,pn) \geq 0$
- L’entropie H est nulle si l’un des événements est certain
- L’entropie H est maximale pour $pi=1/n$
- Le remplacement de $p1,\ldots,pn$ par des moyennes $q1,\ldots,qn$ conduit à une augmentation de l’entropie (convexité de l’entropie).

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

  $H(Y) + H(X|Y)$

- $H(X,Y) \geq H(X)$ ou $H(Y)$
  - $H(X|Y) \leq H(X)$ (égalité ssi indépendance)
- $H(X,Y) \leq H(X) + H(Y) \leq 2.H(X,Y)$
- $I(X;Y) = H(X) - H(X|Y)$

  $= H(Y) - H(Y/X)$

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

- et donc $I(X;Y) \geq 0$ (égalité ssi indépendance)

La relation de ces mesures avec l'entropie est représentée par la figure suivante

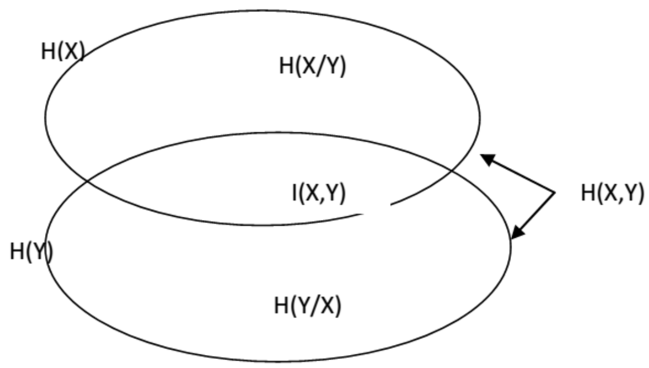

Figure Relation avec l'entropie

4. Conclusion

L'entropie, l'information propre, l'entropie jointe et bien d'autres mesures définies par Shannon, sont les bases de la théorie de l'information. Ce chapitre a donné la formulation mathématique de chacune de ces mesures.