THEORIE

DE L'INFORMATION

jp.guédon

INFO 3

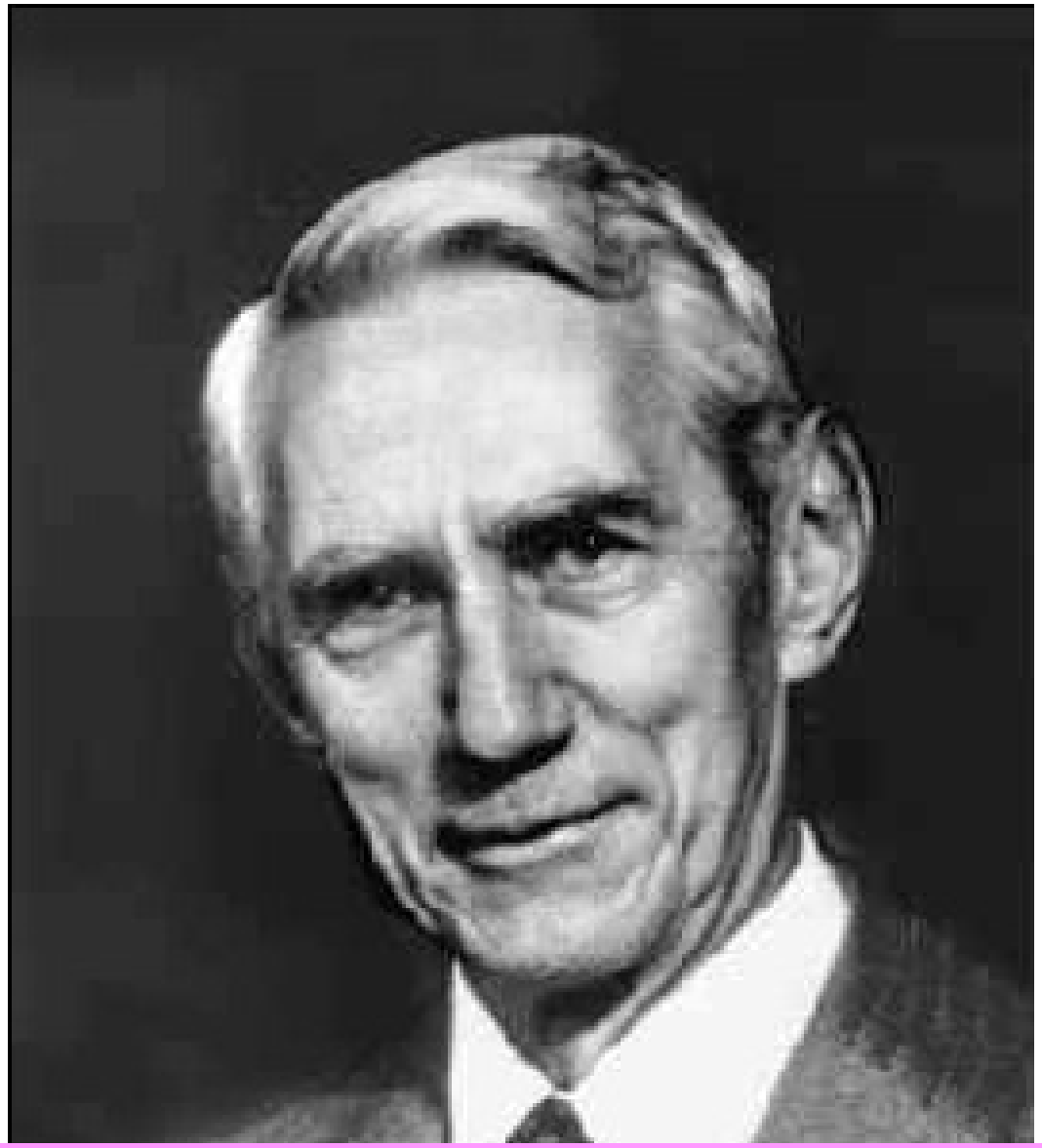
Codages de l'information

Partie 1 l'information

- 1 . entropie
- 2. codage réversible de la source
- 3. information mutuelle

Partie 2 la théorie des codes détecteurs/correcteurs

- 1 Définitions et Quelques exemples
- 2. Codes linéaires
- 3 Codes cycliques



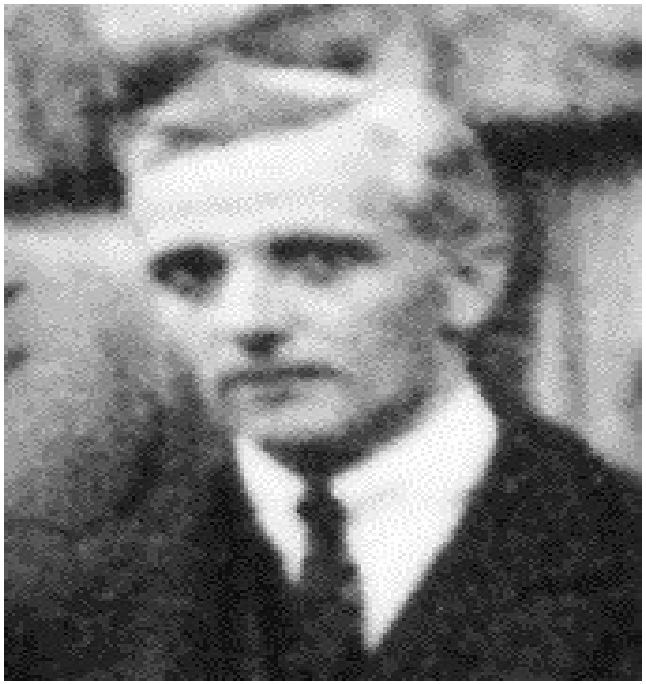
- **Claude Shannon**

Claude Shannon

- **Né le 30 avril 1916, Gaylord, MI, USA**

- **1936: EE Université du Michigan**
- **1940: PhD Massachusetts Institute of Technology (MIT)**
- **1941-1972: Bell Laboratories**
- **Dès 1956: Professeur au MIT**
- **"*A Mathematical Theory of Communication*", 1948: Base de la théorie de l'information**
- **"*Communications in the presence of noise*", 1949: Preuve du théorème d'échantillonage de Nyquist**
- **"*Programming a Computer for Playing Chess*", 1950: Minimax search**
- **Mort le 24 février 2001, Medford, MA, USA**

Léon Brilloin



Léon Brilloin

Léon Nicolas Brillouin, né le 7 août 1889 à Sèvres et mort en 1969 à New York.Il est connu pour ses travaux en mécanique quantique et en physique du solide. Il a notamment travaillé sur la théorie des ondes et la théorie de l'information.

En Août 39, un mois avant la déclaration de guerre à l'Allemagne, Léon Brillouin est nommé, en tant que spécialiste de la propagation des ondes, directeur de la Radiodiffusion nationale. Ayant occupé la fonction quelque temps pendant le régime de Vichy, il démissionne fin 1940 et émigre aux États-Unis.

Léon Brilloin

Léon Nicolas Brillouin émigre aux États-Unis fin 1940.

Il enseigne aux USA dans plusieurs universités (dont celle de Harvard), et travaille chez IBM de 1948 à 1954. Il est élu membre de l'Académie des sciences américaine en 1953.

Il terminera sa carrière et sa vie aux États-Unis.

Il publie en 1959 *Science et théorie de l'information* (version anglaise en 1962) où sont examinées les relations entre ces deux disciplines. Il adopte notamment un point de vue de physicien et fait le lien entre l'entropie informationnelle de Shannon et l'entropie statistique de Boltzmann.

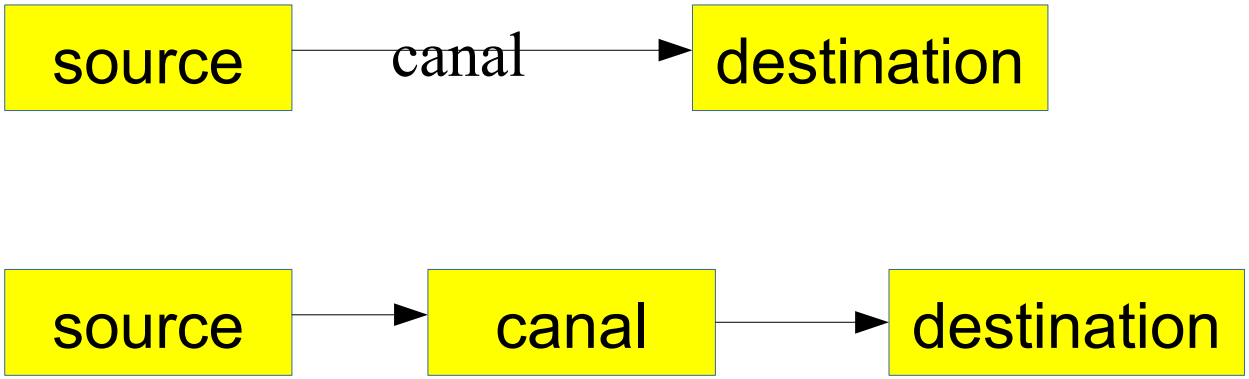
Théorie de l'information

- 1 . information et entropie
- 1.1. définitions
- 1.2 notion d'entropie
- 2. codage reversible de la source
- 2.1 typologie des codes
- 2.2 codage entropique
- 3 . information mutuelle
- 3.1. définitions
- 3.2 le canal de transmission (capacité)
- 3.3 théorèmes de Shannon

Théorie des codes detecteurs/correcteurs

- 1 Définitions et quelques exemples
- 2. Codes linéaires
  - 2.1 Définitions
  - 2.2 Codage / Décodage
  - 2.3 Codes de Hamming
  - 2.4 Autres codes linéaires
- 3 Codes cycliques
  - 3.1 Polynome générateur
  - 3.2 Codes BCH
  - 3.3 Codes RS

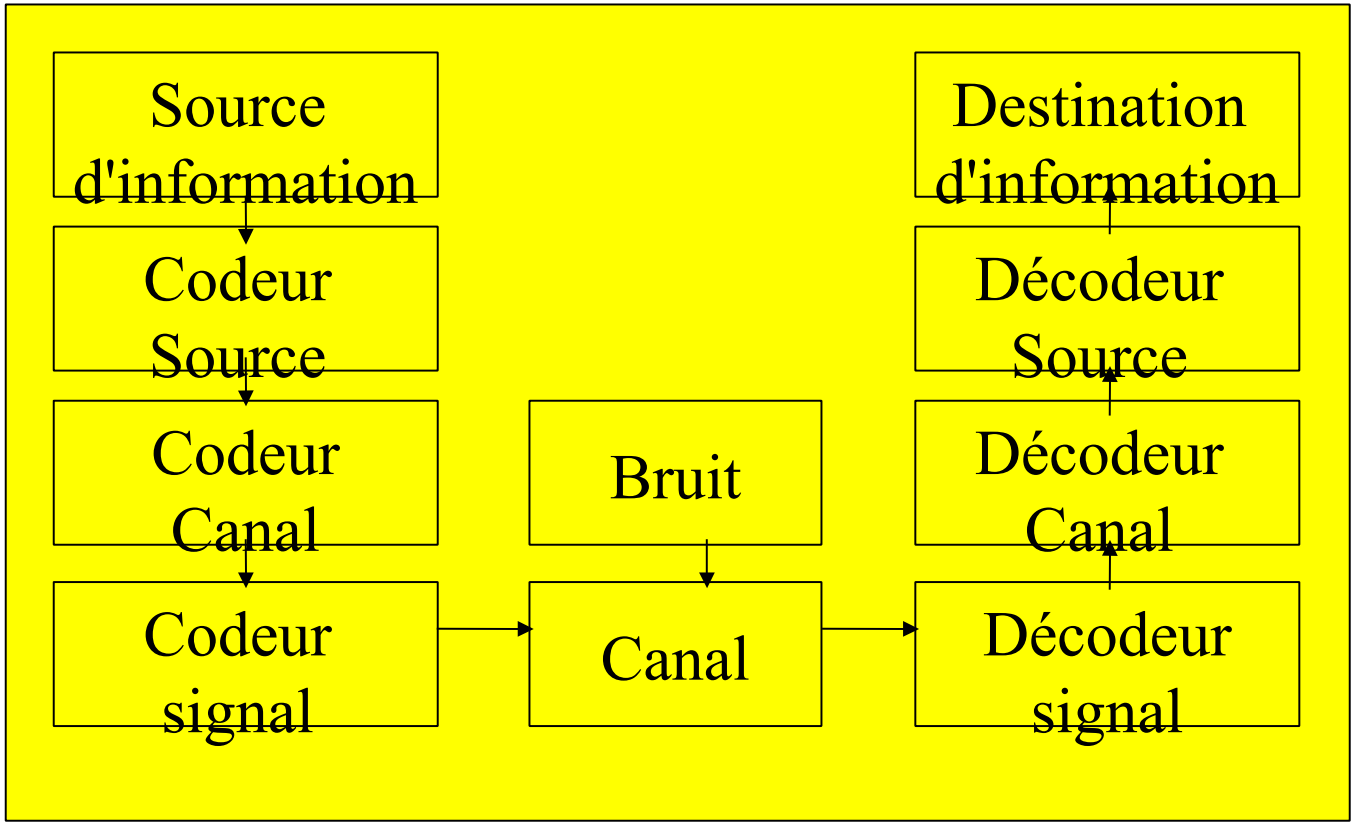
Théorie de l'information



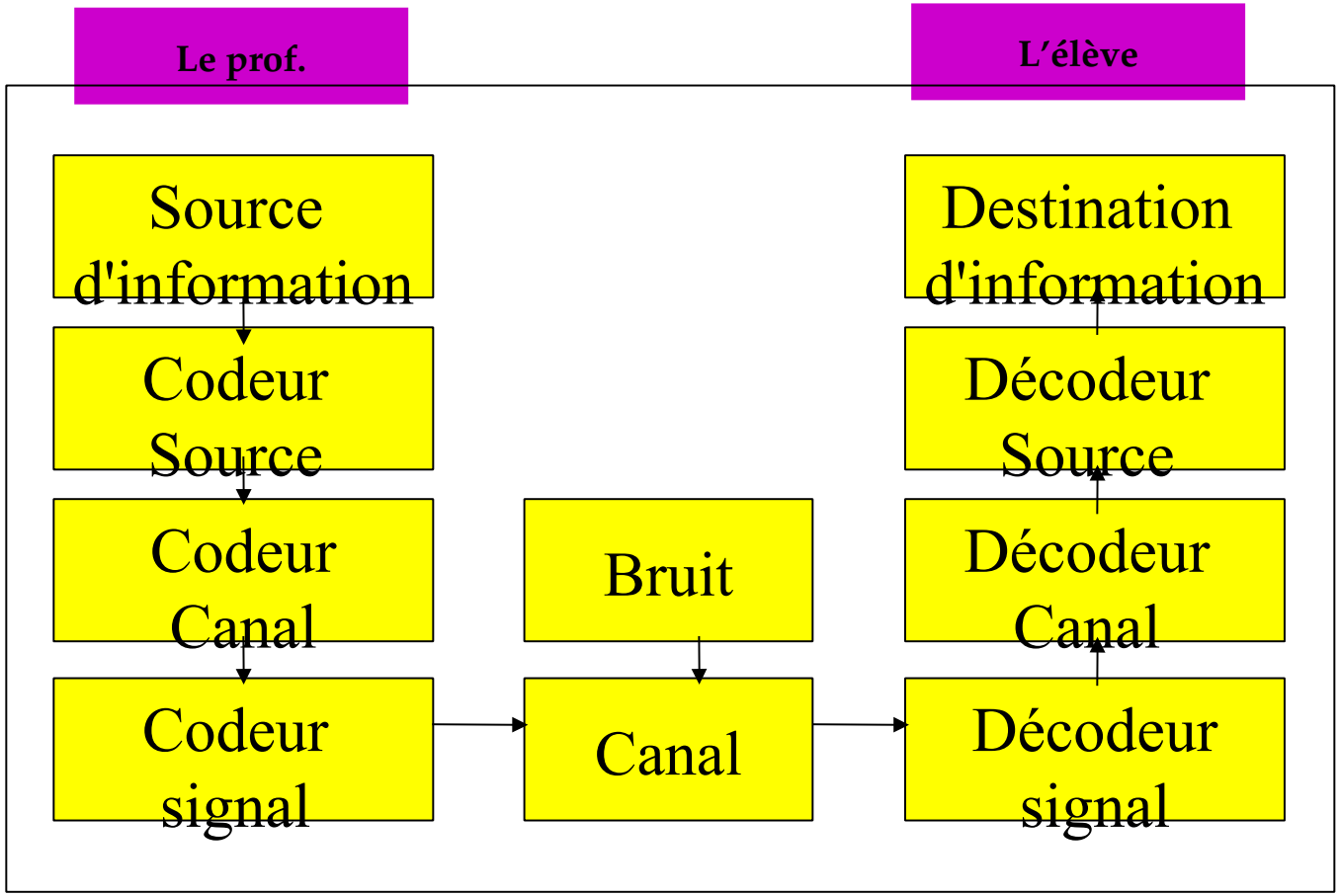
Exemple de canaux :

1 : transmission (réseau)
2 : stockage (disque)

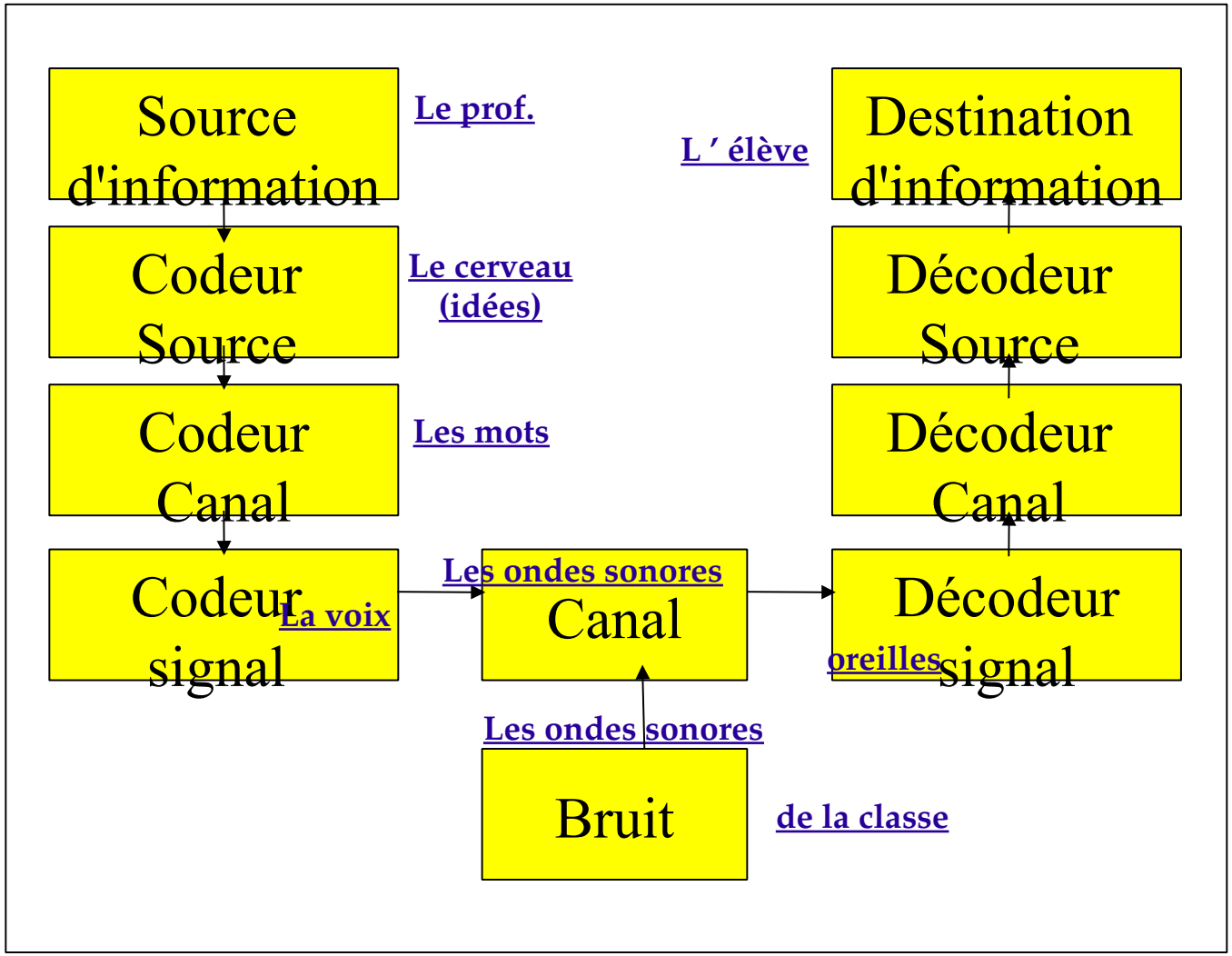
Théorie de l'information



Théorie de l'information



Théorie de l'information



CHAPITRE 1
L'INFORMATION

Introduction [petit Robert]

Information : du latin informatio (1274) : renseignements ou évènements.

Informatique : science de l'information (1962).

Théorie de l'information

exemples d'information :

le journal télé de 20 heures (suite d'images/son, vidéo),

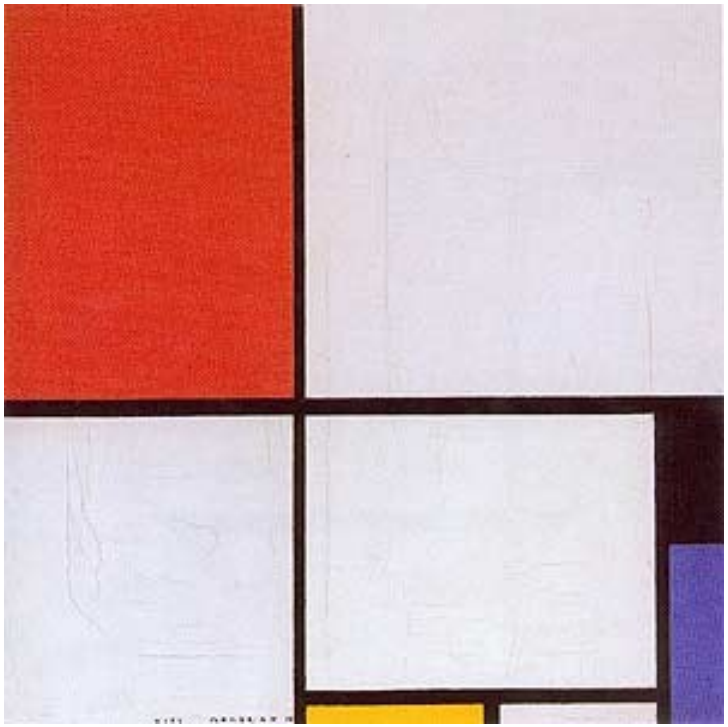
le journal papier (caractères, images fixes),

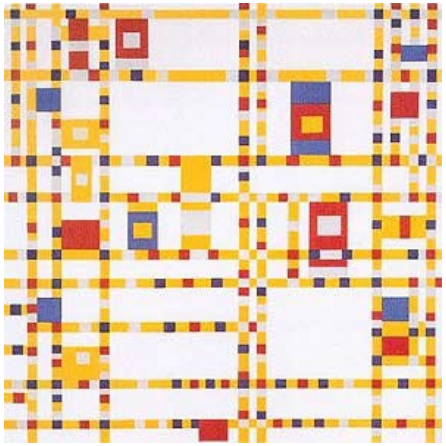
un tableau de Mondrian ou Soulages,

les prévisions météo du lendemain,

les concertos pour violoncelle de Bach

Pierre Soulages





Piet Mondrian

Théorie de l'information

Sur ces exemples d'information :

- Quelle est la quantité d'information contenue dans chaque exemple ? (comment la mesurer ?)

- Quelle est la qualité de chacune de ces informations ? (comment la mesurer ?)

- Est elle vraie? (origines-destinations, perturbations)

- Quelle est sa durée de vie ?

Théorie de l'information

1 L'entropie (la mesure du désordre)

L'entropie est un mot que l'on rencontre un peu partout (thermodynamique, astronomie, etc..).

définition : c'est ce qui caractérise le bazar!

Plus l'ordre règne dans un système fermé et moins on observe du "nouveau".

La mort d'un système intervient lorsqu'il n'y a plus du tout d'agitation (thermique, cosmos, information).

Théorie de l'information : l'entropie

1 L'entropie (la mesure du désordre)

1.1.1 Source d'information, alphabet

Soit une source discrète possédant un alphabet limité A de S symboles différents ai (i=1,...,S) de probabilité d'apparition a priori p(ai).

$$\sum_{i=1,S} p(a_i) = 1$$

Cette probabilité est différente de la probabilité a posteriori qui correspond à la probabilité que le message reçu au récepteur soit correct.

Théorie de l'information : l'entropie

1 L'entropie (la mesure du désordre)

1.1.1 Source d'information, alphabet

Plus la probabilité d'apparition d'un symbole est faible, moins on s'attend à son apparition.

Par exemple, une v.a. d'alphabet A= {0,1} qui suit une loi de Bernoulli B(127/128; 1/128) n'apporte que très peu d'information tant que l'on reçoit le symbole 0.

On obtient donc plus d'information lorsque l'on reçoit un symbole rare.

1 L'entropie (la mesure du désordre)

1.1.2 Auto-information

On définit l'auto-information d'une source discrète par :

$$I(a_i) = -\log_2(p(a_i)) \qquad \textbf{(1.2)}$$

Ce résultat (Shannon 1948) correspond à l'information transmise par le symbole ai .

Théorie de l'information : l'entropie

1 L'entropie (la mesure du désordre)

1.1.2 Auto-information

Dans notre exemple précédent on calcule :

$$I(0) = -\log_2\left(\frac{127}{128}\right) \simeq 1{,}13$$

$$I(1) = -\log_2\left(\frac{1}{128}\right) = 7$$

Remarque importante de calcul :

$$\log_2(p(x)) = \frac{\log(x)}{\log(2)}$$

On remarque donc que le symbole d'apparition très importante ne convoie que très peu d'information.

a contrario le symbole rare est lui très porteur d'information.

Théorie de l'information : l'entropie

1 L'entropie (la mesure du désordre)

1.1.2 Pourquoi un log ?

Comment faire pour classer 2^n choses dans une boite en carton ?

Correction

Théorie de l'information : l'entropie

1 L'entropie (la mesure du désordre)

1.1.2 Pourquoi un log ?

Exemple : pour classer 2^n choses dans une boite en carton la solution la plus rapide consiste à séparer les choses en 2 tas de 2^{n-1} par un carton et recommencer.

Cela permet de retrouver une chose par une étiquette portant le chemin.
La solution la moins rapide consiste à former une liste (1D) de ces choses:

- la recherche demande alors un parcours de la liste

Théorie de l'information : l'entropie

1 L'entropie (la mesure du désordre)

1.1.2 Pourquoi un log ?
En utilisant la base 2, le résultat est exprimé en bits.

(BIT : mot raccourci de BInary digIT trouvé par Tuckey et Shannon)

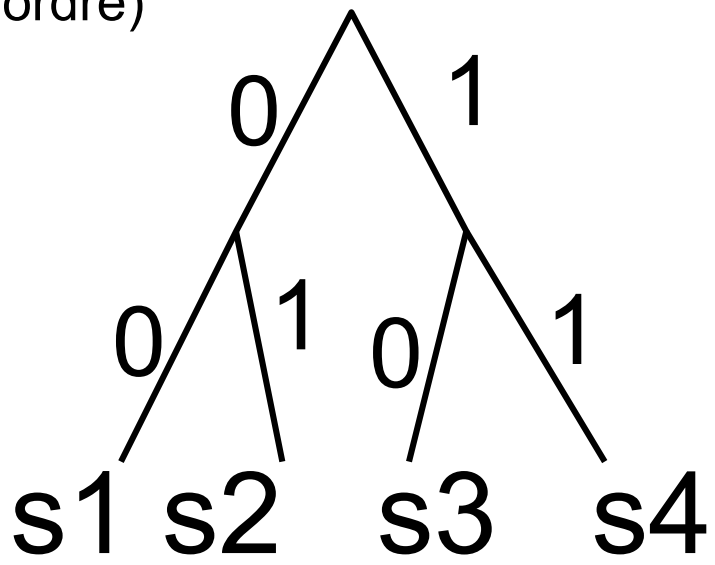
logarithme de base 2 dans la formule : pour coder M symboles différents il suffit de les représenter comme des feuilles dans un arbre binaire.

Théorie de l'information : l'entropie

1 L'entropie (la mesure du désordre)

1.1.2 Pourquoi un log ?

Ex avec 4 symboles



=> L'accès à un élément quelconque se fait alors en descendant à gauche (disons codé par 0) ou à droite (codé alors par 1) à chaque étape.

On a besoin au plus d'un chemin de longueur de $\log_2(S)$ pour identifier un élément unique.

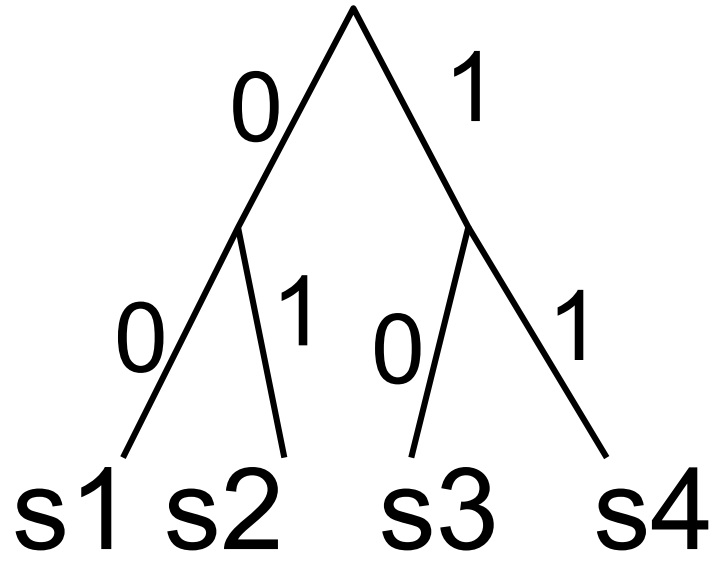
Si l'on transmet ce chemin plutôt que le symbole on ne transmet au plus que $\log_2(S)$ bits.

Théorie de l'information : l'entropie

1 L'entropie (la mesure du désordre)

1.1.2 Pourquoi un log ?

Ex avec 4 symboles



quelle que soit la longueur des symboles s1, s2, s3, s4
il suffit de le remplacer respectivement par 00, 01, 10 et 11

soit log2(4) = 2 bit par symbole

(Attention on n'a pas transmis le dictionnaire dans cette affaire !).

Théorie de l'information : l'entropie

1 L'entropie (la mesure du désordre)

1.1.2 de l'auto-information à l'entropie

Au lieu d'utiliser la mesure par symbole il est plus utile de moyenner sur tous les symboles possibles pour obtenir un chiffre global.
En supposant une source dont tous les symboles sont de probabilité d'apparition indépendants, l'information moyenne par symbole est facile à calculer.
Dans un message de N symboles le symbole i apparait Np(ai) fois en moyenne. L'information totale du message est donc :

$$I_N(A) = N \sum_{i=1}^{S} p(a_i) I(ai) = -N \sum_{i=1}^{S} p(a_i) \log_2(p(a_i)) bit$$

Théorie de l'information : l'entropie

1 L'entropie (la mesure du désordre)

Exercice

Une base de données de bits contient 1Tbit d'information.
Calculer l'auto-information si
1) un seul est à 1 (tous les autres à 0)
2) la moitié des bits sont à 1.

NB: 1Tbit d'information = un message de $N=2^{40}$ symboles

Théorie de l'information : l'entropie

1 L'entropie (la mesure du désordre)

Comme on pouvait s'y attendre l'auto-information est maximale dans le cas ou la probabilité d'apparition du 0 et du 1 sont égales.

Pour comprimer l'information du premier cas de figure (liste de tous les trains qui sont arrivés à l'heure et de l'unique qui a déraillé)

il suffit de savoir localiser l'intrus (l'élément d'information) dans la banalité (absence d'information a posteriori si on sait que les trains arrivent en général à l'heure).
Cela demande en réalité 41 ou 42 bits d'information

Dans le second cas … c'est mission impossible!

Théorie de l'information : l'entropie

1 L'entropie (la mesure du désordre)

1.2. Définition de l'entropie

L'entropie H(A) est la moyenne de $I_N(A)$ sur les N symboles (l'alphabet contient S symboles) :

$$H(A) = E\{I_N(A)\} = -\sum_{i=1,S} p(a_i)\log_2(p(a_i))\, bit/symbole$$

H(A) mesure l'incertitude ou le désordre d'un système.

Théorie de l'information : l'entropie

1 L'entropie (la mesure du désordre)

1.2. Définition de l'entropie

rappel math:

$$H(A)=E\{I_N(A)\}=-\sum_{i=1}^{S} p(a_i)\log_2(p(a_i))\,bit/symbole$$

$$H(A)=\sum_{i=1}^{S} p(a_i)\log_2\left(\frac{1}{p(a_i)}\right)bit/symbole$$

H(A) mesure l'incertitude ou le désordre d'un système.
Plus la valeur de H(A) est petite plus l'ordre règne. (l'arbre se réduit à un tronc sans branche.)

Théorie de l'information : l'entropie

$$H(A)=E\{I_N(A)\}=-\sum_{i=1,S} p(a_i)\log_2(p(a_i))\,bit/symbole$$

Cas extrêmes :

Si on a p(ai)=1 et p(aj≠i) = 0 on obtient *H(A) = 0*. Ce qui veut dire qu'on est toujours sûr de la valeur du prochain symbole : il n'y a plus de surprise possible et donc aucun désordre.

Au contraire, si tous les symboles sont équiprobables cad p(ai)= 1/S on a la source la plus désordonnée (aucun moyen de savoir la valeur du prochain symbole) ce qui conduit à la formule :

$$H(A)=\frac{1}{S}\sum_{i=1}^{S}\log_2(S)=\log_2(S)\,bit/symbole$$

C'est le cas d'un jeu avec un dé à N faces.

Théorie de l'information : l'entropie

1.2. Définition de l'entropie

$$H(A) = E\{I_N(A)\} = -\sum_{i=1,S} p(a_i)\log_2(p(a_i))\, bit/symbole$$

A RETENIR

$$0 \le H(A) \le \log_2(S)\, bit/symbole$$

H= 0 : 1 seul symbole actif
H= log2(S) : chaque symbole à la même probabilité

1.2. Définition de l'entropie

$$H(A) = E\{I_N(A)\} = -\sum_{i=1,S} p(a_i)\log_2(p(a_i))\, bit/symbole$$

Exercice

Une base de données de bits contient 1Tbit d'information.
Calculer l'entropie si
1) un seul est à 1 (tous les autres à 0)
2) la moitié des bits sont à 1.

1.2. Définition de l'entropie

$$H(A)=E\{I_N(A)\}=-\sum_{i=1,S} p(a_i)\log_2(p(a_i))\,bit/symbole$$

1.2.1 Propriétés

1. La valeur de l'entropie est donc toujours bornée entre 0 et log2(S). L'entropie est donc toujours positive.

Exercice

Démontrer la propriété précédente dans le cas d'un alphabet A = {0,1}. En déduire la valeur de H.

démo : soit p(0) = p_0 et p(1)= p_1 les probabilités associées à chaque symbole. On a $p_0 + p_1 = 1$

et $$H(A)=-p_0\log_2(p_0)-p_1\log_2(p_1)\,bit/symbole$$

Théorie de l'information : l'entropie

Exercice
Démontrer la propriété précédente dans le cas d'un alphabet A = {0,1}. En déduire la valeur de H.

démo : soit p(0) = p_0 et p(1)= p_1 les probabilités associées à chaque symbole. On a $p_0 + p_1 = 1$

et $$H(A) = -p_0 \log_2(p_0) - p_1 \log_2(p_1)\, bit/symbole$$

A finir !

Théorie de l'information : l'entropie et la moyenne

$$H(A) = E\{I_N(A)\} = -\sum_{i=1,S} p(a_i)\log_2(p(a_i))\, bit/symbole$$

Variations sur l'entropie

On constitue des messages basés sur l'alphabet constitué des 4 symboles : A={0, 1, 2, 3}.

La probabilité d'apparition de chaque symbole est équiprobable (donc égale à 0,25).

Théorie de l'information : l'entropie et la moyenne

$$H(A) = E\{I_N(A)\} = -\sum_{i=1,S} p(a_i)\log_2(p(a_i))\, bit/symbole$$

Pour le premier message, on place 1000 symboles « 0 » avant 1000 symboles « 1 » avant 1000 symboles 2 avant 1000 symboles « 3 ».

Message = 0000 … 0011111…1122222…22333333…33

Théorie de l'information : l'entropie et la moyenne

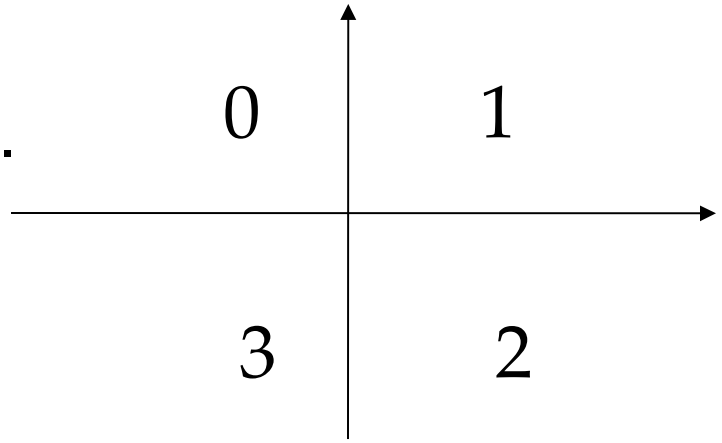
$$H(A) = E\{I_N(A)\} = -\sum_{i=1,S} p(a_i)\log_2(p(a_i))\, bit/symbole$$

Pour le second message, on note chaque jour la température d'une même pièce de la façon suivante :
- moins de 16° on code « 0 »,
- entre 16° et 18° on code « 1 »,
- entre 18° et 20° on code « 2 »
- et enfin au dessus de 20° on code « 3 ».

Théorie de l'information : l'entropie et la moyenne

$$H(A) = E\{I_N(A)\} = -\sum_{i=1,S} p(a_i)\log_2(p(a_i))\, bit/symbole$$

Pour le 3ème message, on partage le plan à partir de l'origine et deux axes orthogonaux.
On attribue à chaque quart de plan la valeur « 0 », « 1 », « 2 », et « 3 » .
On tire au hasard un couple de valeur (x,y) et on code son appartenance au quart de plan.



Théorie de l'information : l'entropie et la moyenne

$$H(A)=E\{I_N(A)\}=-\sum_{i=1,S} p(a_i)\log_2(p(a_i))\,bit/symbole$$

L'entropie est bien entendu la même dans la trois situations.

Cependant, on imagine assez vite que cette source de message ne donne pas du tout la même impression selon l'algorithme utilisé pour produire les messages.

Si on dit maintenant que cette source est envoyée à un receveur, le comportement de celui ci devrait varier pour « s'accorder » sur les propriétés de l'algorithme en même temps que celle de la source.

Théorie de l'information : l'entropie et la moyenne

$$H(A) = E\{I_N(A)\} = -\sum_{i=1,S} p(a_i)\log_2(p(a_i))\, bit/symbole$$

Pour le message 1, le récepteur saura très vite s'adapter. On dit qu'il peut deviner de mieux en mieux la prochaine valeur qui va arriver en ayant vu l'ordre de la suite déjà arrivée. Cette suite est en effet corrélée au maximum.

Théorie de l'information : l'entropie et la moyenne

$$H(A) = E\{I_N(A)\} = -\sum_{i=1,S} p(a_i)\log_2(p(a_i))\, bit/symbole$$

Pour le message 2, le récepteur va également pouvoir corréler très souvent car on peut supposer que la température (comme tout ce qui est naturel) va évoluer doucement.

Par contre on ne peut pas complètement prédire ce qui se passe. La plupart du temps on doit se limiter à ce qui s'est passé les derniers jours pour faire la prévision suivante qui aura une bonne probabilité de se réaliser (par exemple 80%) mais on ne peut pas prévoir les changements.

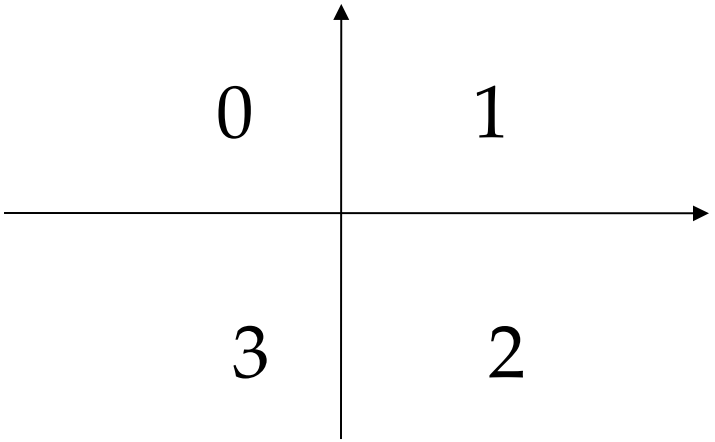
Pour le message 3 ?

Théorie de l'information : l'entropie et la moyenne

$$H(A)=E\{I_N(A)\}=-\sum_{i=1,S} p(a_i)\log_2(p(a_i))\,bit/symbole$$

Pour le message 3, bien entendu on ne peut plus rien prévoir.

Il n'y a aucune corrélation entre une valeur et la valeur suivante.



Théorie de l'information : l'entropie et la moyenne

$$H(A) = E\{I_N(A)\} = -\sum_{i=1,S} p(a_i)\log_2(p(a_i))\, bit/symbole$$

Pour conclure sur cette expérience, on voit que la source est toujours caractérisée de la même façon.

Par contre pour un système comprenant source et récepteur, on peut travailler pour connaître les caractéristiques statistiques de la source

=> cela permet un meilleur décodage voire un codage plus optimal comme dans le cas du MICD qui sera vu dans ce cours.

- **merci**