Théorie de l'information

D. Kohel

04/01/2021

Théorie d'information

Introduction : On traite le problème de la transmission efficace de d'information et sa securité, et surtout les transformation de l'information qui permet d'effectuer ces transmissions.
Les domaines principaux de la théorie d'information sont :

Le codage de l'information

① la compression des données : comment distribuer l'information de manière uniforme et avec concision.

② correction d'erreurs
distribuer et étaler l'information avec redondances (e.g. par repétition) pour proteger

l'information contre des pertes en raison du bruit dans le canal de transmission. [integrité]

La cryptographie

③ securité de l'information: les codages (= chiffrements) qui assurent la confidentialité, l'intégrité (avec preuve, contre un adversaire) et l'authentification des messages et de leurs expediteurs.

Ces trois aspects de la théorie de l'information font partie de la théorie de Shannon, introduit en 1948.

Représentation de l'information

Définitions Un *alphabet* A est un ensemble fini des lettres (ou caractères ou symboles)

Exemple • $\{A, B, \ldots, Z\}$ classique, plus éventuellement des minuscules, ponctuation, espaces, etc.

• $\{0,1\}$ binaire (bits)

• Octets ("bytes" en anglais) $= \{0,1\}^8$

Un *mot* est une suite finie de caractères dans un alphabet. L'ensemble des mots de longueur n sur A s'écrit A^n, et

$$A^* = \bigcup_{n \geq 0} A^n$$

est l'ensemble des mots. Un sous ensemble $M \subseteq A^*$ de mots

(ou de messages) et une langue (une langue formelle ou naturelle).

Remarque. Les messages M d'une langue peut être équipé avec une fonction de probabilité.

Un codage d'un ensemble M de msgs (pas forcement dans A^*) est une application

$$C : M \longrightarrow A^*.$$

Si C est injectif, on dit que C est sans pertes. Un décodage est une application inverse :

$$D : C(M) \longrightarrow M.$$

L'ensemble $C(M)$ s'appellent les mots du code, $C(M) \subseteq A^*$ le code.

Un chiffrement est un codage pas facilement inversible (sans informations supplementaires). Par rapport à un chiffrement, on parle des messages ou des textes clairs de $\mathcal{M}$ et leurs images $c = C(m) \in C(\mathcal{M})$ sont des textes chiffrés. On appelle $\mathcal{C} = C(\mathcal{M})$ l'espace des textes chiffrés.

Exemple. Si $\mathcal{M} = \mathcal{C} = A^n$, où $\mathcal{M}$ correspond à des mot de la langue français (des phrases, etc. de longueur n), on a des messages (= mots, phrases, ..) qui sont plus probable que

d'autres:

$A = \{A, \dots, Z, \_\}$, $n = 8$,

m = JAI SOIF est plus probable que m = AKZMBFXN.

Une fonction de probabilité:

$p : \mathcal{M} = A^n \longrightarrow [0,1]$

donne les fréquences des mots qui apparaissent dans la langue française, tel que $\sum_{m \in \mathcal{M}} p(m) = 1$.

Alors souvant on parle de l'<u>espace des messages</u> pour évoquer que $\mathcal{M}$ est une espace de probabilité (existence de $p : \mathcal{M} \longrightarrow [0,1]$).

Même si $\mathcal{M} = A^n$, on peut avoir de mots $m \in \mathcal{M}$ avec $p(m) = 0$ (e.g. m = AKZMBFXN).

Au contraire, souvant en théorie d'information, on cherche des codages (ou chiffrements) avec la propriété que $C(M)$ est un espace de probabilité uniforme.

Attention. Pas possible si M est fini et $C : M \to A^*$ est sans pertes.

Ex. (Codages) ASII et ISO-8559-1 sont des codages qui envoient des symboles alpha-numériques à des octets $\{0,1\}^8 \subseteq \{0,1\}^*$, qui sont facilement décodés par l'ordinateur.

Théorie de Shannon (1948)

Mathematical theory of communication par Claude Shannon.

- La compression soulève le problème de la mésure quantitative de l'information, et de la redondance de l'information.
- L'integrité de transmission soulève les problèmes de
  - taux de transmission maximal sur un canal avec bruit
  - la construction explicite des codages, avec taux de transmission (ou rendement) donné permettant une

transmission fiable

- La sécurité de l'information soulève le problème de l'<u>information mutuelle</u> (entre l'espace des messages et l'espace des textes chiffrés).

<u>Théorie de probabilité discrète</u>.

Soit X un ensemble dénombrable (ou fini) avec une fonction de probabilité associée :

$$p : X \longrightarrow [0,1]$$

qui vérifie (par définition) :

$$\sum_{x \in X} p(x) = 1.$$

Un tel ensemble est un <u>espace de probabilité</u>.

Défn Un **évènement** est un sous-ensemble d'un espace de probabilité. La probabilité de $E \subseteq X$ est $p(E) = \sum_{x \in E} p(x)$.

Remarque. On a défini une extension de p à $2^X = \{E \subseteq X\}$, où on identifie $X \subseteq 2^X$ par $x \longmapsto \{x\}$. C'est une extension de $p : X \longrightarrow [0,1]$ car $p(\{x\}) = p(x)$.

Remarque. En général (théorie de proba non discrète) on a besoin de distinguer un **mesure** μ sur X et la probabilité

$$p(E) = \int_E d\mu.$$

Dans le cadre discret,

la probabilité est déterminer par $p(\{x\}) = p(x)$ et on remplace l'integrale par une somme.

Ex. Espace de proba uniforme.
Si X est fini, on défini la probabilité uniforme $p : X \to [0,1]$ par $p(x) = 1/|X|$. On a alors

$$p(E) = \frac{|E|}{|X|}.$$

Probabilité conditionnelle

Défn. La proba conditionnelle d'un évènement A, sachant qu'un autre évènement B s'est réalisé, est

$$p(A|B) = \frac{p(A \cap B)}{p(B)}.$$

Lemme. La probabilité cond. sur B est une fonction de proba sur X : $p(\ |B) : X \longrightarrow [0,1]$.

Preuve. Exercice.

Théorème (Bayes)
Si A et B sont des évènements, alors

$$p(B)p(A|B) = p(A)p(B|A).$$

Preuve. Clair (les deux côtés sont égaux à $p(A \cap B)$ par défn).

Indépendence

Défn Deux évènements A et B sont *indépendents* si

$$p(A \cap B) = p(A)p(B).$$

Lemme. Si A et B sont indépendents, alors $p(A|B) = p(A)$ (et $p(B|A) = p(B)$).

Preuve. Exercice.

Variables aléatoires et espérance.

Défn. Une fonction $F : X \longrightarrow \mathbb{R}$ sur X s'appelle une variable aléatoire. (X un espace de proba.)

L'espérance mathématique de F est $E(F) = \sum_{x \in X} F(x)\, p(x)$, un "moyenne" ou somme pondérée par la fonction de probabilité.

Remarque. Dans le cadre non discrète, l'espérance de F est l'integrale $\int_X F(x)\, d\mu(x)$.

Remarque. Une variable aléatoire F induit une fonction de propa sur $Y = F(X) \subseteq \mathbb{R}$: $P_Y : Y \longrightarrow [0,1]$, $y \longmapsto P(F^{-1}(y))$

Exercice. Montrer que

$$E(id_Y) = \sum_{y \in Y} y\, p_Y(y) = E(F)$$

Entropie

Mesure de la quantité d'information d'un espace, introduit par Claude Shannon.

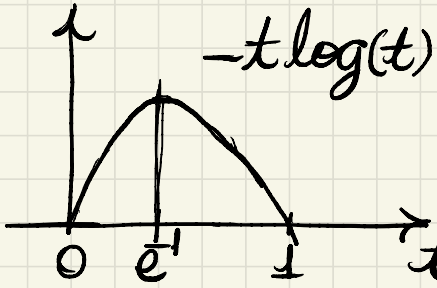
Defn. L'entropie de $X (= (X, p))$ est l'espérance mathématique de la variable aléatoire

$$F(x) = -\log_2(p(x)) :$$

$$H(X) = \sum_{x \in X} \log_2\left(\frac{1}{p(x)}\right) p(x).$$

On s'appelle l'unité d'entropie un bit d'information.

Remarque Si $p(x)=0$, on définit $\log_2\left(\frac{1}{p(x)}\right)p(x)=0$, ce qui est cohérent avec la limite $\lim_{t\to 0} t\log(t)=0$. Graphiquement



Donc, pour $t=p(x)=0$ la définition est cohérente.

Exemples $X=\{0,1\}=\{$"pas beau", "beau"$\}$, valeurs du temps à Marseille, à Paris, etc.

x	$p(x)$
0	0,25
1	0,75

$$H(X)=0{,}25\log_2(4)+0{,}75\log_2(4/3)$$
$$=0{,}50+0{,}75(2-\log_2(3))$$
$$\doteq 0{,}4150<1.$$

L'idée : La réponse à la question : Est-ce qu'il fait beau? porte 0,4150 bits d'information.

$Y = \{0,1\}$ (même ensemble)

$x \in Y$	$p(x)$
0	0,50
1	0,50

= la proba uniforme

$H(Y) = 0{,}50 \log_2(2) + 0{,}50 \log_2(2) = 1.$

Pour la proba uniforme, la réponse à la question porte la quantité maximale d'information (= un bit).

$Z = \{0,1\}$

$x \in Z$	$p(x)$
0	0
1	1

$H(Z) = 0 \log_2(0^1) + 1 \log_2(1)$ (0 → 0^1)

$= 0 + 0 = 0$

Aucune information portée par la réponse.

En général, si X est un espace de proba uniforme ($p(x) = 1/|X|$), alors

$$H(X) = 1/|X| \left(|X| \log_2 |X|\right) = \log_2 |X|.$$