QUATRIEME COLLOQUE SUR LE

TRAITEMENT DU SIGNAL ET SES APPLICATIONS

Nice 7 au 12 mai 1973

THEORIE ALGEBRIQUE DE L'INFORMATION

J.A. VILLE
CIT-ALCATEL
Université PARIS VI, 4 Place Jussieu - 75230 PARIS-CEDEX 05

RESUME

Les problèmes où peut intervenir la théorie de l'information sont ceux de formation et interprétation de messages, et également de transmission de messages ; ils sont connus sous la forme de modulation, démodulation, transfert, filtrage, codage. Interviennent également, en cas de multiplexage, des règles conversationnelles et des règles de stockage. Enfin, couvrant le tout, des règles de fiabilité. Il est essayé de montrer des éléments communs à ces différentes notions.

SUMMARY

I - ENTROPIES DE SHANNON ET RENYI

Dans la définition de l'information d'après SHANNON, on peut faire intervenir la théorie des questionnaires. On s'appuie sur le résultat suivant :
Si un système peut prendre n états avec probabilités $p_1, p_2, \ldots, p_n$, et que l'on ait le droit de poser n'importe quelle question dichotomique pour s'informer de l'état pris par le système, le nombre moyen de questions à poser, en supposant qu'elles soient posées de manière à rendre ce nombre moyen minimum, est égal à

$$(1) \quad H = -\sum_k p_k \log_2 p_k$$

à une unité près par défaut. Ce qui revient à considérer que le système ne prend en pratique que 2^H états.

Si l'on veut se dégager de la théorie des questionnaires, on peut faire appel à une autre interprétation. On part de la formule évidente

$$(2) \quad n = \sum_k \frac{p_k}{p_k}$$

La moyenne harmonique d'une variable x étant

$$x = \frac{1}{\sum \frac{p_k}{x_k}}$$

nous voyons que (2) s'écrit

$$(3) \quad n = \frac{1}{\mathcal{H}_p}$$

ce qui revient à écrire d'une manière générale

$$(4) \quad n = \frac{1}{\mathcal{M}_p}$$

$\mathcal{M}$ représentant une moyenne quelconque. On sait qu'une moyenne quelconque est de la forme

$$\mathcal{M} = g\left[\overline{f(x)}\right]$$

où la barre désigne la moyenne arithmétique et où f et g sont deux fonctions quelconques inverses l'une de l'autre. Ceci conduit à une évaluation de n comme inverse d'une probabilité moyenne

$$(5) \qquad n \sim \frac{1}{g\left[\overline{f(p)}\right]} = \frac{1}{g\left[\sum p_k \, f(p_k)\right]}$$

En choisissant convenablement f(p), on obtient pour log n les différentes formes d'entropie, parmi lesquelles celles de SHANNON et RENYI. Seules sont utilisables en pratique les fonctions f telles que la valeur "pratique" de n donnée par (5) soit inférieure à n.

Si on opère avec l'entropie de RENYI avec f(p) = p nous obtenons

$$n \sim \frac{1}{\sum_k p_k^2} \leqslant 2^{-\sum_k p_k \log_2 p_k}$$

de sorte que cette entropie réduit davantage n que celle de Shannon.

Si l'on transpose le calcul de l'information de Shannon, qui conduit, pour une matrice de probabilités jointes ϖ_{ij} , à

$$(6) \qquad I_S = \sum_{ij} \varpi_{ij} \log \frac{\varpi_{ij}}{p_i \, q_j} \qquad p_i = \sum_j \varpi_{ij} \qquad q_j = \sum_i \varpi_{ij}$$

en utilisant l'entropie de RENYI ci-dessus mentionnée, on trouve

$$(7) \qquad I_R = \log \sum_{ij} \frac{\varpi_{ij}^2}{p_i \, q_j} \geqslant I_S$$

L'information de RENYI étant surestimée ne pourra jamais être atteinte. Il faut noter maintenant que l'expression

$$M = \sum_{ij} \frac{\varpi_{ij}^2}{p_i\, q_j}$$

est une expression rencontrée ailleurs en statistique. C'est une évaluation du rang de la matrice $\varpi = [\varpi_{ij}]$. Si en effet on considère la matrice ρ :

$$\rho_{ih} = \sum_j \frac{\varpi_{ij}\, \varpi_{hj}}{p_i\, q_j}$$

on constate qu'elle a même rang que ϖ, que ses valeurs propres sont réelles, positives, de module au plus égal à 1, et que M est sa trace. Donc M est au plus égal au rang de ϖ_{ij} . Donc

$$I_S \leq I_S \leq \log r \qquad r = \text{rang de } [\varpi_{ij}]$$

Lorsque $I_S = \log r$, la matrice ρ , qui est stochastique, est de plus idempotente. ρ et ϖ sont alors décomposables en r blocs diagonaux qui sont chacun de rang 1, et il est alors possible de transmettre sans altération r messages.

II - EXEMPLES D'EVOLUTION D'ENTROPIE

Soit le système différentiel linéaire

$$(1) \qquad \frac{dx_i}{dt} = \sum_j a_{ij}\, x_j$$

où les a_{ij} sont strictement positifs, à l'exception des éléments diagonaux, qui sont strictement négatifs. Les a_{ij} satisfont les relations

$$(2) \qquad \sum_i a_{ij} = 0$$

qui assurent

$$(3) \qquad \sum_i \frac{dx_i}{dt} = 0$$

Il existe un $k > 0$ tel que la matrice

(4) $\quad k_{ij} + a_{ij}$

soit à éléments strictement positifs. Il existe donc un vecteur propre strictement positif, $\overline{x_i}$, et une valeur propre associée strictement positive λ de la matrice ci-dessus, tels donc que

(5) $\quad \sum_j (k_{ij} + a_{ij})\, \overline{x_j} = \lambda\, \overline{x_i}$

La sommation en i, tenu compte de (2), conduit à

$$k = \lambda$$

et par conséquent à

(6) $\quad \sum_j a_{ij}\, \overline{x_j} = 0$

Revenons au système (1), et partons de valeurs initiales x_i strictement positives, de somme égale à 1. Nous normons le vecteur $\overline{x}$ de manière que $\sum \overline{x_i} = 1$. Etudions la fonction

(7) $\quad F = \sum_i x_i \log \dfrac{x_i}{\overline{x_i}}$

Tenu compte de (3), nous voyons que

(8) $\quad \dfrac{dF}{dt} = \sum_i \dfrac{dx_i}{dt} \log \dfrac{x_i}{\overline{x_i}} = \sum_{ij} a_{ij}\, x_j \log \dfrac{x_i}{\overline{x_i}}$

Considérons maintenant l'expression

(9) $\quad \sum_{ij} a_{ij}\, x_j \log \dfrac{\overline{x_j}}{x_j}$

D'après (2) elle est nulle. Nous pouvons donc écrire

(10) $\quad \dfrac{dF}{dt} = \sum_{ij} a_{ij}\, x_j \log \dfrac{x_i}{\overline{x_i}} \dfrac{\overline{x_j}}{x_j}$

Utilisons maintenant l'inégalité élémentaire

$$\text{Log } u \leq u - 1$$

Nous en déduisons que

(11) $\quad \dfrac{dF}{dt} \geq 0$

Nous venons de démontrer, par un artifice analogue à celui de Boltzmann, que l'entropie des x_i considérées comme des probabilités, va en croissant, en nous repérant par rapport aux $\overline{x_i}$. La croissance de l'entropie est donc une conséquence élémentaire de la linéarisation des équations régissant l'évolution d'un système de probabilités.

Il faut remarquer que le logarithme ne joue pas un rôle essentiel dans la démonstration. Si nous posons en effet

$$G = \sum \frac{x_i^2}{\overline{x_i}}$$

c'est-à-dire si nous faisons intervenir l'entropie de RENYI, on montre d'une façon aussi aisée que G est monotone.

Si on veut démontrer que l'entropie croît, il suffit de linéariser le problème. C'est ce que fait l'hypothèse du chaos moléculaire. Si nous sommes, en face d'un problème de cinétique, face à un système de la forme

$$\frac{dx_i}{dt} = f_i(x_1, x_2, \ldots)$$

dont nous sommes bien embarrassés pour donner la forme parce que nous l'ignorons, un moyen simple de le linéariser est de supposer que les couples $x_i\ x_j$ obéissent à un système linéaire. Nous écrivons donc

$$(12)\quad \frac{d}{dt}\ (x_i\ x_j) = \sum_{st} a_{ij}.st\ (x_s\ x_t)$$

forme que le lecteur admet sans difficulté. Ce système est astreint à être de la forme (1). On somme ensuite par rapport à un des indices, et l'on en déduit (11). Ce mode de raisonnement est assez dangereux. En effet, si le nombre des variables x_i est égal à n, nous avons en (12) un système de n^2 équations différentielles pour n fonctions, qui en général n'admet pas de solution.

III - INFLUENCE DE RESTRICTIONS SUR DES SUITES BOOLEENNES OU POSTIENNES

Si nous prenons comme point de départ qu'une suite de N variables booléennes peut prendre 2^N formes différentes si on n'impose à ces variables aucune contrainte, et que la quantité d'information, dans ces conditions est :

$$(1) \quad I = \mathrm{Log}_2 N$$

nous pourrons définir la quantité d'information que peut apporter une pareille suite, astreinte à la condition

$$(2) \quad F(x_1, x_2, \ldots, x_N) = F(x) = 1$$

où F est une fonction booléenne, comme le logarithme de base 2 du nombre de solutions de l'équation (2).

Supposons par exemple que (2) exprime qu'il n'existe dans la suite $\{x\}$ aucun 1 isolé ni aucun 0 isolé, c'est-à-dire que trois x_n consécutifs ne se présentent jamais sous une des formes

$$(3) \qquad 0\,|\,0$$
$$\qquad\quad |\,0\,|$$

on peut calculer le nombre de solutions de (2). Pour N = 2, c'est 4, pour N = 3 c'est 6. D'une manière générale, ce nombre de solutions est de la forme

$$A \left(\frac{1+\sqrt{5}}{2} \right)^{N-2} + B \left(\frac{1-\sqrt{5}}{2} \right)^{N-2}$$

Nous sommes amenés à évaluer le flux d'informations, pour une variable émise, par la grandeur

$$(4) \quad \log_2 \frac{1+\sqrt{5}}{2} = \log_2 1{,}6 \sim \frac{2}{3}$$

Cela revient à dire que la restriction (3) supprime l'information, en moyenne, pour une variable émise sur 3. Cherchons maintenant à respecter les exclusions (3), et cela par une règle d'émission simple. La plus simple est de ne transmettre que des signaux

doublés. Nous sommes alors certain de ne jamais rencontrer une des configurations (3), et nous ne perdons qu'une variable sur deux, ce qui est honorable, vue la limitation (4).

Comment se concevra un canal qui refuse les configurations (3) ? Si nous restons en binaire, ce canal ne pourra qu'altérer les suites qui présentent ces configurations. Il ne peut en effet "annuler" aucune variable, puisque nous n'avons à notre disposition aucun niveau correspondant à absence de signal.

Supposons que le canal fasse subir aux x_n la transformation

(5) $y_n = \bar{y}_{n-2}\ y_{n-1} + (y_{n-2}\ y_{n-1} + \bar{y}_{n-2}\ \bar{y}_{n-1})\ x_n$

Si la suite $\{x\}$ est "correcte", elle sera transmise sans altération, ce qui est bien. Mais on vérifie que quelle que soit la suite $\{x\}$, la suite $\{y\}$ qui s'en déduit ne présentera <u>jamais</u> les séquences (3), et que par conséquent rien à l'arrivée ne permettra de déceler que le message reçu est faux.

Pour pouvoir exclure certaines configurations par un canal, nous devons opérer en algèbre ternaire, avec les niveaux $\pm$ 1 et 0 (absence de signal). Si par exemple nous excluons les séquences

$$+1\ \ -1\ \ +1$$
$$-1\ \ +1\ \ -1$$

nous pourrons considérer un canal tel que $\{x\}$ devienne $\{y\}$, tel que

$y_n = x_n$ si $\{x_{n-1}, x_n, x_{n+1}\} \neq \{+1, -1, +1\ ;\ -1, +1, -1\}$

$y_n = 0$ dans les autres cas.

Dans une suite de messages, nous obtiendrons ainsi une quantité d'information de la forme

$$\frac{(1+\sqrt{5})^N}{2}$$

N étant le nombre de $\pm$ 1, les règles de transmission étant supposées respectées. La largeur de bande sera caractérisée

par $\frac{2}{3}$ l'énergie par le nombre de $\pm$ 1 .
On peut essayer de choisir une modulation au départ permettant d'utiliser au mieux le canal. Si cette modulation opère dans un intervalle assez long, on peut essayer de transmettre sur les absences de signaux également (puisqu'il n'y a aucune restriction sur les zéros). Mais le décalage est alors compliqué, et on n'est pas gardé contre les erreurs de formation de message.
En effet, un message

+++ - +++

sera reàu comme

+++ 0 +++

et nous ne saurons pas si ce zéro n'était pas un blanc.
Nous sommes amenés ainsi à adopter des règles de modulation telles que celle que nous avions envisagée plus haut, où l'on double les signaux d'information, et les blancs. Une erreur de modulation isolée ne peut ainsi passer inaperçue.

IV - SUR LA NOTION DE RETARD A LA DEMODULATION

Un canal booléen sera défini comme une application de la forme

(1) $\quad y_n = f(x_n, x_{n-1}, x_{n-2})$

en se limitant à une mémoire finie. Même une forme aussi simple pose de difficiles problèmes. Un signal reconstituable sera une suite x telle qu'il existe une inverse permettant de déduire x de y . Ceci amène à considérer l'infinité d'équations

(2) $\forall\, n \quad y_n = f(x_n, x_{n-1}, x_{n-2})$

et à résoudre en x. La question de l'unicité n'a de sens que si on se donne la forme de la fonction inverse. Il ne sert de rien de chercher à résoudre (1) en x_n. Pour le cas extrêmement simple où

(3) $\quad y_n = x_n\, x_{n-1}$

on trouverait que l'unicité de x_n est liée à $x_{n-1} = 1$.

Mais on peut se donner par exemple comme solution :

$$(4) \quad x_n = y_n + y_{n+1}$$

ce qui impose à la suite x de ne pas présenter de 1 isolé. L'essentiel est de remarquer que bien que le canal ne semble pas retarder le signal, d'après (3), il le retarde en réalité, puisqu'une formule telle que (4) ne donne x_n qu'une fois connu y_{n+1}.

D'une manière générale, le retard réel n'apparaît pas sur une formule telle que (1). Il faut faire une convention de modulation, et liée à une règle de restitution du signal. La capacité n'apparaît pas non plus directement. Elle est liée au retard que l'on s'autorise dans la reconstitution du signal. Elle est naturellement d'autant plus grande que le retard admis est plus grand.

Nous voyons ainsi que pour des règles algébriques simples de transfert, les notions de modulation, démodulation, retard, capacité ne sont pas indépendantes, ni définies isolément.

V - EXEMPLE DE MODULATION LUTTANT CONTRE UN BRUIT DE DEPLACEMENT EN TEMPS

Le filtrage peut être utile pour adapter un message à un canal. Seront correctement modulés les signaux non altérés par le filtre. Il peut également intervenir pour éliminer un bruit. Le filtre joue alors le rôle de décodeur. Mais encore faut-il qu'il y ait eu codage, c'est-à-dire modulation convenable. Restant dans le domaine booléen, considérons une suite x, et une suite ε de signaux de bruit. Nous considérerons le cas où le bruit affecte la phase, c'est-à-dire par exemple où

$$y_n = \bar{\varepsilon}_n x_n + \varepsilon_n x_{n-1}$$

L'impulsion de bruit décale le signal. Si nous voulons pouvoir

éliminer le bruit, il nous faut évidemment pouvoir retrouver tous les x_n. Une modulation qui consiste à tripler le signal permet de repérer les décalages isolés. Supposons en effet que 0 et 1 soient transmis par :

0 : 0 1 1

1 : 1 0 0

Il n'apparaît jamais, dans cette modulation, de séquence de 4. Or le bruit de phase considéré, lorsqu'il n'est pas inoffensif, fait apparaître de pareilles séquences. Le triplement ordinaire ne se prête pas à un filtrage homogène, parce qu'il faut conserver la synchronisation.

VI - MODULATION EN ALGEBRE TERNAIRE

Soit un message dont nous voulons doubler tous les signaux. Si nous ne faisons pas intervenir de synchronisation, nous aurons besoin d'un message dont un signal sur 2 est nul. Ceci peut s'obtenir par la condition de récurrence

$$(1) \quad x_{n-1}^2 + x_n^2 - 2\, x_{n-1}^2\, x_n^2 = 1 \qquad (x_n^3 = x_n)$$

Tous les signaux de rang pair seront nuls, ou tous les signaux de rang impair, sans que nous distinguions les deux cas, puisque nous n'avons pas d'origine des temps.

Le modulateur pourra être conçu comme fournissant

$$(2) \quad y_n = x_n + (1 - x_n^2)\, x_{n-1}$$

Les signaux non nuls sont doublés. y_n ne prend que les valeurs ± 1 .

Le démodulateur pourra être de la forme

$$(3) \quad \xi_n = y_n(1 - \xi_{n-1}^2)$$

Nous voyons que dans ces conditions, $\{\xi_n\}$ est de même nature que $\{x_n\}$, avec un décalage éventuel constant, sans que dans

(2) ni (3) nous ayons fait intervenir d'origine des temps. Nous avons ainsi un exemple de modulation démodulation permettant d'utiliser un canal qui refuse les séquences + - + et - + - , mais il a été nécessaire, pour une spécification complète, de faire intervenir une algèbre ternaire.

Si on remarque que les règles conversationnelles séparent des portions de message par des zéros, nous voyons que le modulateur ne pourra, si on ne veut pas perdre d'information, que remplacer des zéros par des $\pm$ 1 . Ceci souligne l'importance de l'étude des systèmes d'équations récurrentes postiennes, qui semble avoir été négligée dans la littérature.