Cliver des Cartes Auto-Organisatrices, une voie pour modéliser la Perception/Cognition?

Splitting Self Organizing Maps, a way of modelling perception/cognition?

Michel Collobert
France Telecom R&D

2 avenue Pierre Marzin 22300 Lannion
michel.collobert@orange-ftgroup.com

Résumé

En nous inspirant des résultats récents de la neurobiologie de la perception, nous présentons dans ce papier un travail de réflexion sur ce que peut être un système où la perception et la décision (ie RF & IA) sont structurellement liées.
Contrairement à un grand nombre de modèles de réseaux de neurones formels, un objet perçu ou une connaissance n'est pas dans notre modèle représenté par l'activité finale "de sortie" d'un seul neurone du réseau mais par la *configuration d'activité de ces neurones répartis en plusieurs sous ensembles à l'instar des aires cérébrales.*
Ce modèle présente des caractéristiques prometteuses. Il est capable intrinsèquement de gérer la fusion sensorielle. Il permet de s'affranchir de la centralité du symbole. Par son aptitude à la gestion des "données manquantes", il fournit aussi un lien entre perception et cognition.

Mots Clef

Bio inspiré, cartes topographiques, cognition, perception, fusion sensorielle.

Abstract

In this paper some global brain activity paradigms (specialization of cerebral areas, topographical mapping, non-linear dynamics) are used to design a new kind of artificial perceptual system based mainly on multiple Kohonen self-organizing maps (SOM). In contrast to most artificial neural network models, a perceived object or knowledge is not represented by a single neuron's final activity of a net but by the whole neurons' set configuration activity.
The new model presented here has promising properties. It manages sensory modalities fusion and missing data retrieval; the provided linkage between perception and cognition may lead to a new "embodied system" paradigm.

Keywords

Bio-inspired, Kohonen Topographical maps, sensory fusion.

1 Introduction

La façon dont le cerveau fonctionne exactement est encore un sujet de controverse dans le champ de la neurobiologie. Cependant, grâce à l'apport des nouvelles techniques de neuro-imagerie, de nouveaux paradigmes émergent ou sont confirmés. Par exemple, bien que certains travaux montrent le rôle de certains neurones isolés [1], il est maintenant évident que, à un instant donné, une partie significative des neurones faisant partie d'une multitude d'aires cérébrales est activée simultanément (figure 1); le traitement, la représentation, la mémorisation de l'information y est hautement distribuée. Seule une compétition locale au sein d'une aire ou même d'une "sous-aire" cérébrale existe.

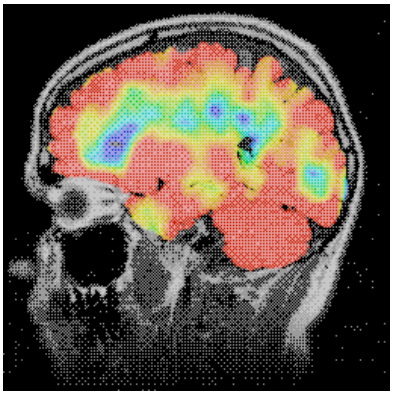

Figure 1: vue sagittale de l'activité corticale.

En dépit de cela, peut-être par le fait que dans notre conscience humaine nous ramenons tout à des mots, i.e. des "symboles", nous avons naturellement tendance à concevoir nos algorithmes pour obtenir un (ou quelques) symboles en "sortie". Même dans le domaine des réseaux neuroniques où l'on peut dire que l'information est distribuée comme les perceptrons multi - couches [2] ou les Cartes auto-organisatrices de Kohonen [3] le processus est similaire. Un seul (ou quelques) neurone du réseau est considéré comme activé à la fin d'un processus de calcul de façon que nous puissions, *nous humains,* comprendre le résultat.

Quelques chercheurs comme Rodney Brooks [4] suggèrent que les limitations actuelles que nous rencontrons en "milieu ouvert" aussi bien dans le domaine de l'Intelligence Artificielle (IA) que dans celui de la Reconnaisance des Formes (RF), spécialement la vision par ordinateur, proviendraient du fait que nous oublions un concept fondamental dans la conception de nos modèles actuels. Le nouveau paradigme d'organisation que nous présentons dans ce papier cherche à combler une petite part du fossé existant avec les "réalisations du vivant" en s'inspirant justement des nouvelles données de la neuro-biologie. Notre but n'est pas de simuler au plus près un système biologique, mais d'essayer de dégager les paradigmes fondamentaux utilisables dans un système "incarné" dans le silicium.

Les 2 paradigmes principaux issus de la neurobiologie que nous mettons en avant dans ce papier sont les multiples aires spécialisées d'un cerveau, et leurs propriétés topographiques (rétinotopie, tonotopie, somatotopie, …). Cela nous conduit à représenter une perception ou une connaissance, non comme l'activité finale d'un seul neurone, mais comme la **configuration d'activité** d'un ensemble de neurones appartenant à des "cartes" spécialisées. Un troisième est la diffusion.

2 Description du modèle

Les deux principes inspirés de la neurobiologie essentiels à notre modèle sont donc :

- La spécialisation fonctionnelle des aires cérébrales : le cerveau humain, entre autres, est divisé en plusieurs aires, chacune traitant de préférence un domaine spécifique de la perception ou de l'action ; nous avons des aires pour la vision, d'autres pour la parole, pour le toucher etc… Chaque aire est elle-même subdivisée en plusieurs sous aires. Par exemple, pour la vision, des aires nommées V1, V2, V3, V4, MT, chacune dédiée au traitement des contours, de la couleur, du mouvement, de la position dans l'espace etc… Même dans l'architecture cognitive d'insectes comme les abeilles, leurs "mini-cerveaux" consistent en un réseau complexe de modules interconnectés (neurones dédiés, neuropiles, centres d'intégration d'ordre supérieur) [5].

- Les propriétés topographiques de chacune de ces aires : une partie significative de ces aires ont des propriétés topographiques respectant à la fois la physique et la statistique des entrées. Ainsi, 2 fréquences lumineuses proches seront traitées sur des endroits proches de l'aire cérébrale auditive, même chose pour les odeurs [6], ou les mouvements. Cette projection est aussi en quelque sorte "non-linéaire", c'est-à-dire que plus le système nerveux doit discriminer de façon fine entre des entrées, plus il utilise de neurones pour ce faire.

Pour simuler ces deux principes "in silico", nous avons utilisé dans un premier temps de Multiples Cartes Auto-Organisatrices de Kohonen (ou KSOM : Kohonen self organizing Maps). Leurs propriétés reflétant les propriétés topographiques du cerveau sont connues [3]. Et nous allons montrer que contrairement à l'usage qui en est couramment fait, leur multiplication (correspondant aux multiples aires cérébrales) peut apporter, malgré quelques inconvénients finalement mineurs, des propriétés inattendues.

2.1 Principe de représentation :

Pour présenter notre modèle à partir d'un exemple simple, il nous faut décrire succinctement l'application d'origine. Celle-ci, à visées télécoms (compression de flux vidéo), consiste à segmenter les images du flux vidéo en zones homogènes au sens d'un certain critère, puis à apparier au mieux ces zones entre 2 images consécutives.

La segmentation de chaque image est basée sur une KSOM monodimensionnelle d'une dizaine de neurones. En indexant chaque couleur de l'espace (R, V, B) par cette carte, on obtient par regroupement des pixels indexés un certain nombre (environ 300) de zones.

Chaque zone est étiquetée par une suite d'attributs :
- sa couleur indexée correspondante ;
- sa surface (i.e. son nombre de pixels) ;
- la position de son centre de gravité en X et Y dans l'image ;
- sa hauteur, sa largeur et le rapport entre les deux (indice de forme).
- sa vitesse.

Ce sont ces zones ainsi étiquetées que le système doit suivre d'une image à la suivante.

L'usage habituel des KSOM, que nous avons testé de prime abord, veut que pour notre application, l'on projette l'ensemble des zones représentées par leurs vecteurs d'attributs sur une carte unique 2D dans une phase d'apprentissage. La phase de reconnaissance consistant alors à trouver le neurone de cette carte présentant la distance minimum avec le vecteur d'entrée à reconnaître.

Dans notre application, nous avons eu la volonté de moduler l'influence de chaque attribut dans le calcul de cette distance car l'apparition ou la disparition de zones sur les bords de l'image posait un problème de variation trop rapide sur certains attributs (surface, hauteur, largeur). Cela nous a conduit à multiplier le nombre de cartes de la façon suivante :
À chaque attribut, nous avons fait correspondre une carte de type KSOM unidimensionnelle (16 neurones) et pour chaque objet la valeur de l'attribut correspondant est indexée par cette carte.

Les figures 2a et 2b montrent les activations (dans un souci de clarté uniquement pour les 3 premières composantes) correspondantes du système pour 2 objets différents :

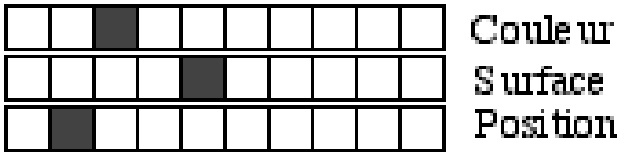

Fig 2a) Configuration d'activité pour l'objet A

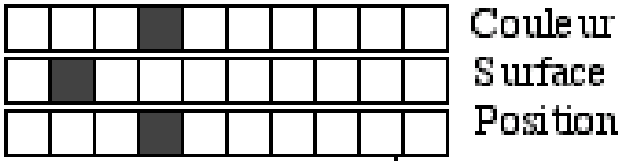

Fig 2b) Configuration d'activité pour l'objet B

Ainsi chaque objet est représenté par une configuration d'activité de l'ensemble des différentes cartes impliquées.

Comme une qualité des KSOM est leur non linéarité, on peut être sûr que la quantification obtenue est optimale pour discriminer entre les objets en utilisant cet attribut unique. Par exemple si l'ensemble des objets à identifier / discriminer est composé de visages, la KSOM associée à l'attribut de couleur est composée de neurones répondant en majorité chacun à une valeur précise de la teinte de la peau.

2.2 Principe de mémorisation :

Pour mémoriser ces configurations, nous avons utilisé dans notre implémentation "in silico" (i.e. un ordinateur) une sorte de structure mimant l'hippocampe [1], en créant pour chaque configuration un neurone connecté à chaque neurone actif des KSOMS. Cela revient d'une certaine façon à de la mémorisation "par cœur", domaine où les ordinateurs excellent.

Nous pouvons noter aussi que contrairement aux cartes de Kohonen classiques, le nombre d'objets mémorisables croît de façon géométrique avec le nombre de cartes utilisées.

2.3 Principe de reconnaissance :

Le cerveau est connu pour son caractère chaotique [7][8]. Il peut être vu comme un système complexe d'oscillateurs couplés, qui passe d'une configuration mémorisée à une autre en l'absence de stimulus externe. C'est ce que montrent les travaux concernant la déprivation sensorielle. A un instant donné, des nouvelles entrées sensorielles forcent le système à se stabiliser dans un état précédemment mémorisé.
Cela est cohérent avec certains neurobiologistes qui affirment que le cerveau essaye de construire une cohérence avec le monde extérieur [9] et les chercheurs en 'vie artificielle' qui parlent de 'dynamically coherent coupling' [10].

C'est ce que dans notre modèle nous appelons la « reconnaissance/perception » et que nous réalisons en utilisant une troisième propriété des neurones biologiques : la diffusion d'activité entre neurones (qui, dans les modèles de réseaux neuroniques, a été utilisée principalement pour l'apprentissage).

Reprenant notre exemple basique, deux cas peuvent être distingués pour la reconnaissance:

- Soit la configuration issue des entrées sensorielles est strictement identique à une des configurations mémorisées. Alors le problème est trivial et résolu.

- Dans le cas contraire, un appariement est nécessaire avec la configuration mémorisée la plus 'proche'. Par exemple supposons que l'on veuille apparier la configuration d'activation de la figure 3 avec une des deux (l'objet A ou l'objet B) décrites figure 2.

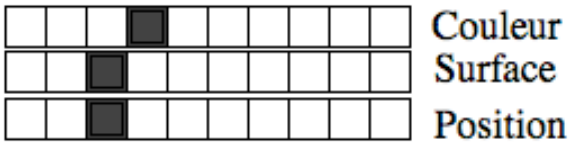

Fig 3) Configuration d'activité pour l'objet à reconnaître

C'est là qu'une notion de distance semble nécessaire. Mais comment comparer des distances entre caractéristiques différentes (par exemple couleur et surface) ? Nous nous retrouvons devant le problème classique de la « bonne » normalisation de chaque caractéristique. C'est là que le caractère non linéaire des cartes de type Kohonen entre en jeu. En première approximation, la quantification que réalisent ces réseaux fait en sorte que chaque neurone, que chaque case de la carte donc, représente une variation égale sur l'espace des entrées. Si l'on prend la précaution d'utiliser des cartes de taille égale, une variation d'une case sur chaque carte sera donc équivalente au point de vue statistique quelle que soit la caractéristique utilisée.

En reprenant notre exemple, si nous utilisons donc en même temps le principe de diffusion sur chaque caractéristique comme indiqué sur la figure 4,

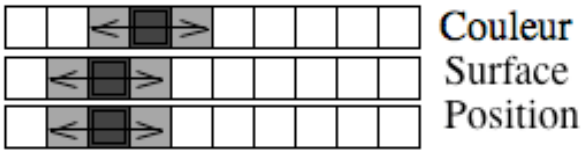

fig 4 : Exemple d'une première étape de diffusion

et cherchons à apparier la configuration d'activité obtenue avec celle de l'objet A ou celle de l'objet B de la figure 2, on voit par surimposition des patterns que c'est l'objet B qui rentre le mieux en « résonance ». Le système s'est donc stabilisé dans la configuration d'activité mémorisée la plus proche. Nous pouvons alors dire que le système a « reconnu » les signaux d'entrée.

Evidemment, si à cette étape, plus d'un candidat « résonne », il est nécessaire de faire entrer un jeu un autre attribut jusqu'à ce qu'un seul pattern gagnant reste.

C'est une caractéristique importante de notre modèle : il n'est pas forcément nécessaire de calculer ou d'acquérir toutes les caractéristiques pour arriver à une perception d'un objet mémorisé.

3. Résultats

Nous avons appliqué ce modèle dans l'application décrite plus haut. Pour nous permettre de gérer par exemple les effets de bords sur l'image, un processeur classique (i.e. un programme) assisté par un « micro système expert » mémorise les configurations, fait les comparaisons et pilote le fonctionnement de la partie multi-cartes suivant le schéma de la figure 5.

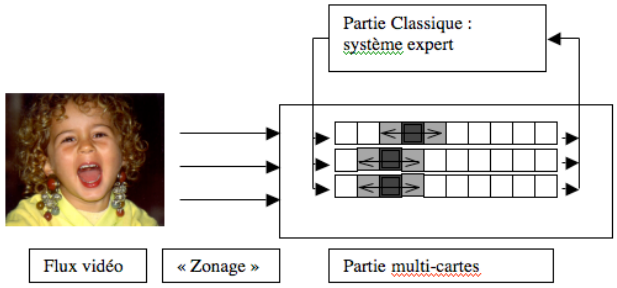

fig 5 : Structure de notre application

Le «micro système expert » sert à injecter de la connaissance a priori dans le *fonctionnement* du « processeur neuronal » avec des règles toutes simples du type :

> Si (« zone recherchée » touche « bord image »)
> Alors (faire diffuser d'abord la carte « nombre de pixels »)

(Cette règle indique simplement qu'une zone touchant le bord de l'image appartient peut-être à un objet qui entre ou sort du champ de la caméra. Donc sa surface est susceptible de varier très rapidement).

Parmi les 300 zones générées à chaque image par la quantification sur la couleur, environ 80% sont suivies d'une image sur l'autre sur une séquence de test d'environ 10 secondes comportant un travelling. A comparer au 50% de zones suivies avec un réseau de kohonen unique, nous estimons ces résultats très encourageants au vu de la segmentation très primaire opérée par la quantification vectorielle uniquement sur la couleur.

4. Discussion

On peut se demander pourquoi utiliser cette sorte de modèle dégénéré de carte de Kohonen et perdre ainsi certaines qualités intéressantes de cette réduction de dimension de l'espace d'entrée (en particulier la gestion de la corrélation entre les différents attributs).

Nous venons de voir dans l'implémentation précédente que nous pouvons injecter de la connaissance *a priori* pour moduler le fonctionnement de chaque carte. Le système est aussi rapide et souple d'utilisation. Mais surtout d'autres avantages se dessinent :

4.1. Fusion sensorielle

De par leur mode de construction, il n'y a pas a priori de problème pour mélanger des cartes issues de modalités sensorielles différentes. Même pour une seule modalité sensorielle, chaque carte est en fait indépendante de la modalité concernée et ne dépend que la statistique des entrées de la caractéristique correspondante. Donc il n'y a pas de problème pour mixer les cartes obtenues pour deux ou plusieurs modalités différentes. Et c'est aussi ce qui œ passe dans les systèmes nerveux biologiques : toutes les modalités sensorielles sont actives en même temps et participent à la configuration globale de la perception de l'organisme vivant.

Par exemple, notre modèle peut être utilisé pour faire de la reconnaissance de parole en utilisant conjointement le son et l'image des lèvres en mouvement. D'un coté des caractéristiques dynamiques issues d'une caméra comme dans [12] peuvent être utilisées pour la partie image. De l'autre coté, partie son, on peut s'inspirer des nombreux travaux [13] qui ont été fait dans le domaine par la communauté connexioniste.

4.2 Mémorisation de prototype d'objets

La mémorisation, cette fois ci, non d'un objet particulier à un instant donné, mais de son prototype peu être effectué par apprentissage au cours du temps de la statistique de ses représentations. Plus une valeur d'un attribut est observée, plus le neurone correspondant est actif. Il n'y a plus de fonctionnement « tout ou rien » Nous pouvons ainsi obtenir une représentation du type de la figure 6.

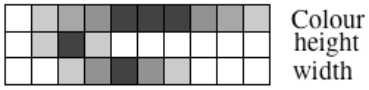

Figure 6. Exemple de configuration d'activité pour un prototype.

4.3 Classification contextuelle.

L'utilisation de multiples KSOMs fournit une façon aisée pour la classification contextuelle. Dans notre modèle, les classes utiles pour un contexte particulier sont vues comme des combinaisons possibles de classes 'primaires' fournies par les KSOMs.

4.4. Gestion des données manquantes

Comme nous l'avons montré plus haut, il n'est pas nécessaire de balayer l'ensemble des cartes pour faire marcher (résonner) notre système. Comme seuls sont mémorisés des états globaux d'activation, une complémentation des données manquantes peut être faite à chaque étape du processus. Cette qualité peut être très utile pour :

- Résoudre des ambiguïtés. Par exemple en vision artificielle, la gestion des occultations partielles est un problème récurrent, généralement résolu de façon ad-hoc.

  Si l'on prend soin d'imiter le système visuel des primates [14] en sur-multipliant les caractéristiques de l'objet ou d'un sous ensemble de l'objet (par exemple la tête par rapport au corps humain), nous sommes sûr que (sauf en cas d'occultation sévère que notre propre système visuel a du mal à résoudre) le système trouvera suffisamment de cartes se stabilisant dans une configuration permettant de distinguer un objet particulier des autres configurations mémorisées.

- Notre modèle permet de s'affranchir de la centralité du symbole, les symboles pouvant être vus simplement, par exemple, comme une configuration d'activité de cartes phonologiques associée à la configuration d'autres cartes représentant le reste d'une perception particulière (i.e. la signification du symbole). Dans notre modèle, le symbole n'est pas la meilleure façon de résoudre les problèmes. Néanmoins, il peut être vu comme le principal moyen de communication (à travers le langage) entre deux systèmes qui ont des structures de perception similaires. La transmission d'un symbole à un récepteur induisant par complémentation un état interne global similaire à l'état global de l'émetteur. Bien sûr, l' « information » (pas au sens de Shannon) ne peut être transmise que si le récepteur a une structure suffisamment voisine du récepteur.

- Simuler l'inférence logique. C'est une version du problème de l'occultation qui est particulière intéressante car notre modèle conduit ainsi à un lien pertinent pour un congrès comme RFIA qui veut lier

les domaines de la Reconnaissance des Formes et l'Intelligence Artificielle.

Si l'on observe que l'on peut remplacer dans le problème de l'occultation les caractéristiques visibles par les prémices d'une règle d'inférence et les données manquantes (obtenues par complémentation) par les conséquences de cette même règle, on obtient un système similaire à un système expert. En fait dans notre modèle, comme il est dit pour le fonctionnement neurobiologique du cerveau [11] nous pouvons aussi dire que « la perception est décision », les deux fonctions y étant indissociablement liées.

Bibliographie

[1] Quian Quiroga R., Reddy L., Kreiman G., Koch C. Fried I., "Invariant visual representation by single neurons in the human brain", *Nature* Vol 435/23, 2005, pp 1102-1107.

[2] Zhang G.P. Neural Networks for Classification: A Survey, *IEEE Trans on Systems, Man and Cybernetics* - Part C, vol 30, n°4, nov 2000, PP451-462.

[3] Kohonen T., Self-Organizing Maps, Springer, Berlin, 1995.

[4] Brooks R., The relationship between matter and life, *Nature* (2001), vol 409.

[5] Giurfa M. The amazing mini-brain: lessons from a honey bee. *Bee World*, 84(1), p5-18

[6] Guerrieri F., Schubert M., Sandoz J-C, Giurfa M. "Perceptual and neural olfactory similarity in honeybees". *PLoS Biology* 3(4), e60.

[7] Faure P., Korn H., Is there chaos in the brain? I. Concepts of non-linear dynamics and methods of investigation. *C. R. Acad. Sci. Paris*, Ser. III 324 (2001) 773-793

[8] Babloyantz, A., Salazar, J.M., & Nicolis, C. "Evidence of chaotic dynamics of brain activity during the sleep cycle", *Physics Letters*, 111A, 152-156 (1985)

[9] Berthoz A. *The brain's sense of movement*. Harvard University Press, 2000. *Le sens du mouvement* , Odile Jacob

[10] Almeida e Costa F., Rocha L., 'Embodied and Situated Cognition' *Artificial Life Volume 11*, Numbers 1-2, January 2005, pp. 5-12(8) MIT Press

[11] Berthoz A., *"La décision"*, Odile Jacob, 2003, p10.

[12] Michael S. Gray, Javier R. Movellan, Terrence J. Sejnowski, "*Dynamic features for visual speechreading: A systematic comparison*", Advances in Neural Information Processing Systems, 1997[13] Christiansen M.H., Chater N., "*Connectionist natural language processing. The state of art*". Cognitive Science 23, 417-437; (1999)

[14] Tanaka K, Orban G. "*Discontinuités dans l'image rétinienne : segmentation et analyse des formes planes".* Séminaire Collège de France, Mai 2007.

[15] Delorme A. « *Traitement visuel rapide de scènes naturelles chez le singe, l'homme et la machine* » Thése Paris 2000.