



RESEAUX DE NEURONES ET FUSION DE DONNEES

V.Schmidlin^{‡†}, G.Favier[‡], J.P.Stromboni[‡]
et B.Tomasini[†]

[‡] I3S-CNRS/Université de Nice-Sophia Antipolis
Bât SPI n°4 - 250, rue. A. Einstein
06560 Valbonne

[†] SYSTELIA Technologies
6, Bd Maréchal Leclerc
83320 Carqueiranne

Résumé: L'objet de cet article est de présenter quelques solutions neuronales pour la fusion de données, et plus particulièrement pour la poursuite multicible multicapteur. Après avoir passé en revue les différentes architectures neuronales existantes, on présente les solutions neuronales utilisables pour la fusion de données multicapteur, en vue de réaliser la détection et le pistage de cibles.

Abstract: The aim of this paper is to present some neural solutions for data fusion, and more particularly for a multiple target tracking purpose. After a review of the different existing neural architectures, we present the neural solutions available for multisensor data fusion, with target detection and tracking purposes.

1. INTRODUCTION

La fusion de données constitue, aujourd'hui, un maillon important de la chaîne de traitement numérique de l'information, qui consiste à passer d'un ensemble d'observations à une prise de décision. L'enjeu est d'améliorer cette prise de décision en fusionnant des informations à la fois hétérogènes, incertaines, incomplètes et redondantes, issues de plusieurs capteurs de types différents, ou du même type mais réglés sur des plages de fonctionnement différentes.

L'objet du présent article est d'effectuer une revue des architectures neuronales dédiées à la fusion de données. Dans le paragraphe 2, on décrit brièvement les principales architectures de réseaux neuronaux, et on indique quelques domaines d'application de ces réseaux. Le paragraphe 3 est consacré à une application particulière des réseaux de neurones: la fusion de données radar, en vue de la détection et du pistage de cibles.

2. RESEAUX NEURONAUX

Si l'on oublie le paradigme biologique, les réseaux de neurones formels, ou réseaux neuronaux (RN), constituent une catégorie de filtres non-linéaires adaptatifs, dont le traitement est hautement parallèle. Il s'agit en principe de résoudre un problème d'optimisation à partir d'un traitement élémentaire identique au niveau de n cellules (neurones formels) interconnectées. Le principe collectif du traitement (connexionnisme) prédispose ces réseaux aux tâches de perception fortement dépendantes d'un contexte. L'engouement actuel pour les RN résulte de la

conjonction de trois caractéristiques principales: simplicité de la mise en oeuvre de chaque cellule, parallélisme massif des calculs et capacité d'adaptation.

On distingue différentes familles de RN, selon la nature du traitement cellulaire, la topologie des connexions, et le principe d'adaptation (algorithme d'apprentissage) utilisé pour la mise à jour des poids des connexions:

- Le traitement cellulaire est constitué d'une somme pondérée des entrées de la cellule (les pondérations étant appelées poids synaptiques), appliquée à une fonction de décision f , appelée aussi fonction d'activation. Cette fonction peut être dérivable (comme la sigmoïde pour les réseaux multicouche [Rum] et le réseau analogique de Hopfield [Hop]), ou non dérivable (comme la non linéarité de type tout ou rien). Les entrées et la sortie de chaque neurone peuvent être binaires ou analogiques.

La sortie du neurone peut aussi être calculée en utilisant des probabilités dépendant des neurones voisins, comme dans la machine de Boltzmann [Fre].

- La structure du réseau est généralement régulière (connexion totale, en bandes, en couches). Les réseaux peuvent être récurrents (on dit aussi bouclés) ou directs, et les connexions totales ou partielles, avec des possibilités de combinaison. Le réseau de Hopfield et la machine de Boltzmann présentent une connexion totale récurrente: chaque cellule est reliée à toutes les autres. Par contre, le réseau multicouche illustre une connexion directe. Les cellules sont organisées en couches totalement connectées: couches d'entrée, cachée(s), et de sortie. Quant au réseau de



Kohonen, il allie une connexion totale et directe des entrées aux cellules qui sont organisées en une couche, et une connexion partielle récurrente des cellules de la couche.

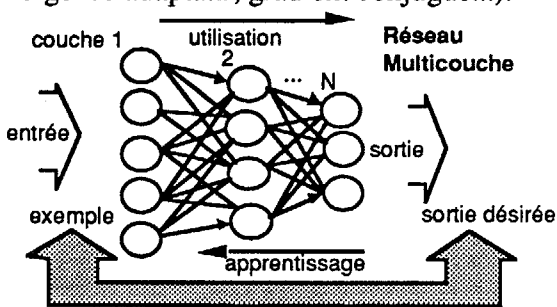
- La troisième caractéristique des réseaux est constituée par leur mécanisme d'apprentissage.

Ainsi, le réseau multicouche qui est certainement le plus utilisé dans les applications de traitement du signal, est doté d'un apprentissage supervisé. On lui présente d'abord un ensemble de couples (*entrée, sortie à associer*), appelé ensemble d'apprentissage. Ces entrées sont propagées de la couche d'entrée à la couche de sortie du réseau. Puis, l'apprentissage consiste à ajuster les poids synaptiques de façon à minimiser, pour ces exemples, l'écart entre la sortie désirée et la sortie effective du réseau. Cette minimisation est en général réalisée à l'aide d'une méthode du type gradient, comme par exemple l'algorithme de rétropropagation du gradient de l'erreur (RPGE). Ce calcul est conditionné par la dérivabilité de la non linéarité du neurone.

Une fois la phase d'apprentissage terminée, le réseau peut être utilisé avec des entrées ne faisant pas partie de l'ensemble d'apprentissage; on parle alors de capacité de généralisation. Cette situation est typique des problèmes de classification.

Plusieurs problèmes se posent lors de l'utilisation d'un réseau multicouche:

- dimensionnement du réseau, c'est à dire choix du nombre de couches cachées et du nombre de neurones par couche; une solution incrémentale peut être utilisée pour résoudre ce problème [Alp];
- détermination de la base d'exemples pour l'apprentissage;
- accélération de la convergence, grâce à des modifications de l'algorithme RPGE (taux de convergence adaptatif, gradient conjugué...).



- Figure 1 -

D'autres réseaux sont munis d'un apprentissage non supervisé qui les entraîne à reconnaître les prototypes de la base des exemples. Ainsi, la loi de Hebb relie les poids à la corrélation des états des neurones connectés. Dans le réseau de Kohonen, l'apprentissage est compétitif. Seules les cellules les plus sensibles à une entrée sont adaptées, dans le sens d'une augmentation de sensibilité.

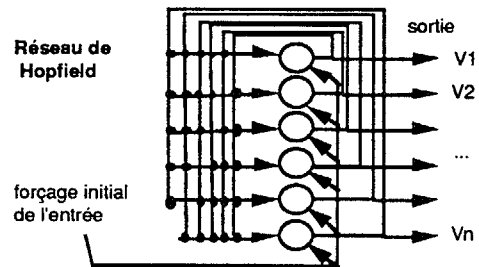
Bien entendu, une question clef avec les RN récurrents est relative à leur convergence vers une solution stable, des preuves de convergence faisant encore défaut pour de nombreuses architectures de réseau. Les principaux outils d'analyse des RN sont fournis par la mécanique statistique et la théorie des

systemes dynamiques.

Ainsi par exemple, dans le cas des mémoires associatives bidirectionnelles et du réseau de Hopfield, l'existence d'états stables peut être montrée en faisant appel au concept de fonction de Lyapounov, appelée aussi fonction énergie, définie comme:

$$E = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij} \times V_j \times V_i$$

Ce type de RN a été utilisé par [Hop] pour résoudre des problèmes d'optimisation, et en particulier le problème classique du voyageur de commerce. L'idée est de ramener le problème à la minimisation d'une fonction énergie, différentes contraintes pouvant être introduites pour améliorer la convergence (symétrie des a_{ij} , nullité des poids diagonaux a_{ij} , ajout de termes dans la fonction à minimiser...). Les poids a_{ij} sont alors calculés par différenciation de la fonction énergie. L'utilisation du réseau consiste à forcer l'état initial des cellules et à laisser évoluer la structure bouclée vers un état d'équilibre qui fournit le résultat du traitement neuronal.



- Figure 2 -

Des réseaux "complexes", où coexistent des cellules différentes, et où les lois d'apprentissage peuvent devenir algorithmiques, ont également été proposés. Ainsi, le réseau ART [Fre] utilise une résonance réglable entre deux couches de cellules entièrement connectées pour reconnaître une forme apprise.

Voici à présent un bref panorama des applications privilégiées des réseaux neuronaux, la classification apparaissant en tête du hit parade de ces applications:

- le traitement des images, en ce qui concerne la mise en correspondance stéréo, l'extraction de caractéristiques, mais aussi la compression ou la restauration d'images,
- le traitement de la parole, aussi bien pour la reconnaissance que pour la synthèse, ou encore pour l'identification de locuteur,
- la reconnaissance de caractères,
- l'automatique et le traitement du signal, avec en particulier l'identification et la commande de systèmes non-linéaires, l'annulation d'échos, la réduction de bruit, ou encore la compression de données,
- la robotique, pour les aspects vision artificielle, calcul de trajectoire, planification de tâche et commande de mouvement,
- le biomédical, dans des buts de diagnostic.

Enfin, le domaine de la Défense fournit de nombreuses applications des réseaux neuronaux: filtrage vrais plots/faux plots, détection et pistage de cibles, classification de signaux radar et sonar, séparation de sources.

3. FUSION DE DONNEES

La mise au point d'un système de veille ou de poursuite performant nécessite l'utilisation de capteurs tout à la fois:

- nombreux, pour pouvoir effectuer d'éventuels recoupements entre les informations déduites de ces capteurs,

- différents, pour tirer parti de leur complémentarité.

Se pose alors tout naturellement le problème de la gestion de toutes ces informations, à la fois complémentaires, donc redondantes, incomplètes, voire contradictoires, et hétérogènes. Enfin, la fusion (ou encore l'agrégation) de toutes ces informations doit s'effectuer très rapidement, les contraintes temps réel étant très fortes. Dans le cadre de la fusion de données appliquée au pistage multicible multicapteur, on peut dégager les deux problèmes principaux suivants:

- la mise en correspondance de ces données avec les connaissances déjà acquises (fusion temporelle)
- la prise en compte globale des données reçues localement et au même instant par les différents capteurs (fusion de capteurs).

3.1 Le pistage

3.1.1 Par mesure de divergence

A l'instant courant t , on dispose d'une base de données constituée de M pistes, et on reçoit un ensemble de N plots à associer aux M pistes pré-existantes. Un réseau neuronal peut être utilisé pour calculer un degré de similitude entre les cibles déjà détectées (pistes i) et les cibles détectées à l'instant t (plots j). Pour cela, il suffit de définir une mesure de distance $\Phi(i,j)$ entre la distribution de probabilité d'une piste i et celle d'un plot j [Bro]. Lorsque cette distance est minimale et inférieure à un certain seuil τ , cela signifie que la piste i se retrouve à l'instant t dans l'espace d'observation, où elle a été détectée par l'intermédiaire du plot j . On a donc réussi à effectuer un suivi temporel de la cible. Si la distance minimale $\Phi(i,j)$ est supérieure au seuil τ , cela signifie qu'à l'instant t , on ne détecte aucune cible susceptible de correspondre à la piste i : on a donc "perdu" la cible.

3.1.2 Par optimisation

Si l'on considère que le problème du pistage revient à associer des plots à des pistes, tout en minimisant de façon globale l'ensemble des distances plot/piste (sur toutes les associations plot/piste possibles), alors on peut se ramener à un problème d'optimisation, similaire à celui du "voyageur de commerce", résolu par un réseau de type Hopfield [Hop], [Her]. Néanmoins, ce type d'approche présente un certain nombre de désavantages:

- le réseau converge en général vers un minimum local de la fonction énergie;
- la solution obtenue ne tient compte d'aucune connaissance a priori sur le problème, dans la mesure où la possibilité d'associer un plot à une piste très éloignée de ce plot, sera envisagée au même titre que

l'association avec une piste très proche.

Notre approche [Sch] a donc été basée sur l'introduction d'une probabilité d'association:

$$\alpha_{xi} = P(\text{associer plot } i/\text{piste } x)$$

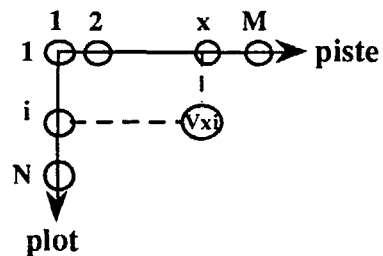
calculée à partir de la distance séparant le plot i de l'estimée de la piste x (à l'aide d'un filtre de Kalman),

$$\text{par la relation: } \alpha_{xi} = \exp(-d_{xi}^2 / 0,5 d_{\max}^2)$$

où d_{\max} est la distance au-delà de laquelle on estime qu'il est pratiquement impossible d'associer le plot i à la piste x . Ceci va permettre d'écartier dès le départ les associations plot/piste irréalistes, et d'initialiser le réseau assez près de la solution, donc d'y converger plus rapidement et plus sûrement. La fonction énergie est alors choisie de la façon suivante:

$$E_{\text{obj}} = \frac{1}{N} \sum_{i=1}^N \prod_{x=1}^M (1 - \alpha_{xi} V_{xi})$$

et la structure neuronale sur laquelle on effectue la minimisation de cette fonction, est un réseau totalement interconnecté, organisé sous la forme d'un tableau:



- Figure 3 -

U_{xi} est l'entrée du neurone placé en xi -ième position, et V_{xi} représente la valeur de sortie de ce neurone. En fin de convergence:

- $\{ V_{xi}=1 \text{ ie on associe le plot } i \text{ à la piste } x$
 - $\{ V_{xi}=0 \text{ ie on n'associe pas le plot } i \text{ à la piste } x$
- Cette fonction énergie aura tendance à faire tendre vers 1 les V_{xi} pour lesquels α_{xi} est grand, et vers 0 ceux qui ont une probabilité d'association faible.

La minimisation de cette fonction est assurée dès lors

que l'on a [Tak]: $\frac{dU_{xi}}{dt} = - \frac{\partial E_{\text{obj}}}{\partial V_{xi}}$ et

$V_{xi} = f(U_{xi})$ avec f telle que: $\frac{dV_{xi}}{dU_{xi}} \geq 0$.

La méthode de calcul est la suivante:

- On dispose de M pistes. Un filtrage de Kalman permet d'estimer le plot attendu pour chaque piste x .
- On reçoit N plots. On calcule l'ensemble des distances plot i /estimée piste x , soit $\{d_{xi}\}$
- On en déduit les probabilités d'association plot i /piste x , soit $\{\alpha_{xi}\}$
- On initialise les neurones d'un réseau totalement interconnecté de dimension $N \times M$, par $U_{xi} = \alpha_{xi}$
- On remet à jour séquentiellement (ligne par ligne) les $\{U_{xi}\}$ par la relation:

$$\Delta U_{xi} = - \frac{\partial E_{\text{obj}}}{\partial V_{xi}} = \frac{\alpha_{xi}}{N} \prod_{y=1, y \neq x}^M (1 - \alpha_{yi} V_{yi})$$

de façon à garantir la minimisation de la fonction énergie



• puis on calcule les $\{V_{xi}\}$ à l'aide de la formule:

$$V_{xi} = \text{sgm}[U_{xi} - 0,5 \text{Max}_{j,j \neq i}(U_{xj}) - 0,5 \text{Max}_{y,y \neq x}(U_{yi})]$$

de façon à garantir: $\frac{dV_{xi}}{dU_{xi}} \geq 0$, et pour inhiber la présence simultanée de deux "1" sur une même ligne ou une même colonne, avec $\text{sgm}(x) = 1/(1+e^{-kx})$.

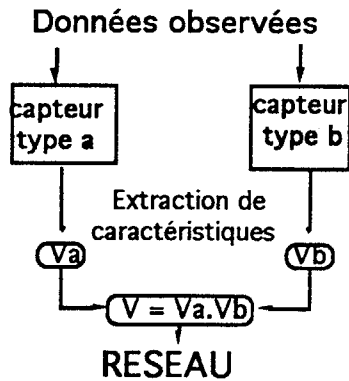
3.1.3 Par reconnaissance de trajectoires

Si l'on considère qu'une trajectoire a une forme déterminée par la cinématique qui la régit, on peut apprendre à un RN à reconnaître la forme de ces trajectoires et à les anticiper. L'apprentissage est alors supervisé, le but étant, pour le réseau, de mémoriser une "bibliothèque" de pistes (appelées piste-prototypes). Ainsi, à chaque piste qu'on lui présente en entrée, le réseau est capable de trouver de quelle piste-prototype cette piste s'approche le plus; il peut ainsi déterminer la position attendue la plus probable pour le tour d'antenne suivant, et effectuer alors l'association plot/piste correspondante [Sch].

3.2 La fusion multicapteur

3.2.1 Fusion centralisée

A l'instant t, des capteurs de type identique ou de types différents, observant une même région de l'espace, fournissent un flot d'informations à l'utilisateur [Ruc], [Pot].



- Figure 4 -

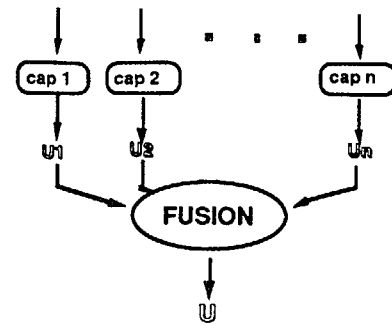
Ces capteurs ont pour but de fournir sur le phénomène observé, des caractéristiques qui vont permettre une bonne reconnaissance de la situation. Il faut tout d'abord réduire le nombre de ces caractéristiques (en gardant les plus discriminantes) de façon à diminuer le temps de calcul. On concatène ensuite les caractéristiques restantes, sous la forme d'un vecteur V, qui est mis à l'entrée d'un réseau multicouche. Ce réseau est auparavant entraîné par l'algorithme de rétropropagation, à reconnaître les caractéristiques qui correspondent à un certain nombre d'hypothèses (dans le cas de la détection de cibles, deux hypothèses: H0=absence de cible et H1=présence de cible), et fournit donc en sortie, la réponse adéquate.

3.2.2 Fusion décentralisée

Ici, il n'y a pas de distinction entre le cas où les capteurs sont de type identique et celui où ils sont de

types différents, puisque les décisions sont prises localement, indépendamment les unes des autres. Le schéma du système est le suivant:

Données observées



- Figure 5 -

U_i est la décision locale prise par le capteur i, et U est la décision globale prise au niveau du centre de fusion: $U=1$ décide l'hypothèse H1: "présence de cible" $U=-1$ décide l'hypothèse H0: "absence de cible"

Dans le cas où les capteurs et le centre de fusion sont soumis à des variations imprévisibles de l'environnement qu'ils observent (par exemple, des changements dans les probabilités a priori des hypothèses de présence ou d'absence de cibles, ou des variations dans le rapport Signal/Bruit des capteurs), il est utile de faire appel à un apprentissage, afin de permettre aux neurones d'adapter leurs poids et leur seuil en fonction de ces variations [Kam]. Ainsi, chaque capteur i doit, à chaque itération t, évaluer les probabilités a priori des hypothèses H0 et H1. De la même façon, au niveau global, le centre de fusion réajuste ses paramètres, et estime les probabilités de fausse-alarme et de non-détection de chaque capteur, afin d'en adapter les caractéristiques et d'assurer ainsi qu'il prenne une décision minimisant la probabilité d'erreur de détection.

4. CONCLUSION

Les réseaux neuronaux suggèrent des réponses aux problèmes posés par la fusion de données. Ainsi, le problème d'association plot/piste peut être résolu par le calcul, en utilisant comme support, une structure neuronale, ou par la minimisation d'une fonction énergie confiée à un réseau de Hopfield. D'autre part, dans un cadre de fusion multicapteur, on peut envisager d'entraîner un réseau multicouche aux tâches de reconnaissance de situation, telles que la détection de pistes à partir d'un vecteur regroupant les caractéristiques du contexte, ou l'adaptation à des variations de l'environnement. Il apparaît d'ores et déjà que les approches neuromimétiques ouvrent dans ce domaine, un champ de possibilités et d'investigations important et prometteur. Toutefois, le problème de la fusion pris dans sa globalité, c'est-à-dire depuis les capteurs jusqu'à la prise de décision, devra certainement être abordé en mariant les méthodes de Traitement du Signal avec les méthodes connexionnistes et les techniques d'Intelligence Artificielle: on parle alors de systèmes hybrides [Fav]. (références bibliographiques fournies sur demande).