

APPROCHE NEURONALE DE LA DETECTION DECENTRALISEE

**Emmanuel DUFLOS, Marie-Paule BOYER,
Yves COQUARD**

I.S.E.N., Département des Signaux et Systèmes,
41 Bd Vauban 59046 LILLE

RÉSUMÉ

Le problème de la détection décentralisée résulte de la volonté de décomposer l'opération de détection en deux niveaux de traitement de façon à réduire la quantité d'information transmise au sein des systèmes. Ces systèmes comprennent un réseau de capteurs suivis par une étape de traitement local et d'un traitement central. Des solutions basées sur des approches conventionnelles probabilistes, qui s'avèrent coûteuses en calculs, ont d'abord été étudiées. Nous proposons une approche neuronale de la détection décentralisée. L'objectif de cette étude est de montrer l'intérêt de l'intégration des méthodes connexionnistes au sein des structures de détection décentralisée en montrant qu'il est possible d'apprendre la relation qui lie les rapports de vraisemblance locaux au rapport de vraisemblance centralisé.

ABSTRACT

The decentralized detection problem appears when various sensors are used simultaneously. In the centralized vision all the information arriving at each sensor is sent to the fusion center, while in the decentralized a local processing is realized before a fusion processing. Most studies on the design of decentralized detection systems consider conventional probabilistic structures which require a lot of calculations. The aim of this paper is to study the possibility to use a neural approach to design a decentralized detection system. We show that it is possible to learn the relation that exists between the local likelihood ratios and the global likelihood ratio which is the optimum centralized system.

1. INTRODUCTION

Les systèmes de détection ont souvent à traiter de grandes quantités d'information provenant non pas d'une source unique mais d'un ensemble de sources. De nouvelles structures de détection décentralisée sont donc apparues ces dernières années qui comprennent, associé à chaque source d'information un premier traitement local suivi d'un traitement central de fusion. Le problème est donc de déterminer les traitements locaux et la loi de fusion qui optimisent les performances de détection par rapport à un critère donné. Nous proposons dans cet article d'aborder le problème de la détection décentralisée par le biais d'une approche neuronale. L'objectif est d'apprendre à l'opérateur central à mettre en commun les informations qui lui parviennent des différents capteurs de façon à ce que le comportement du système décentralisé ainsi conçu se rapproche le plus possible du système optimal centralisé classique.

Il est important de noter que la solution au problème de la détection à structure libre préconisée par la théorie classique de la décision repose sur le postulat de centralisation de l'information qui suppose d'emblée que l'information ainsi que le

traitement qui lui est appliqué soient regroupés en un même lieu. La mise en oeuvre de cette structure nécessite une grande capacité de transport d'information et donc s'adapte mal aux conditions dans lesquelles les systèmes de détection sont amenés à opérer. Le problème de la détection décentralisée résulte donc de la volonté de décomposer l'opération de détection en deux niveaux de traitement de façon à réduire la quantité d'information transmise au sein des systèmes. Une alternative à la structure centralisée peut être proposée sous la forme d'une structure imposée où le traitement est décomposé en plusieurs étapes, par exemple où chaque source élabore un résumé de son observation, que l'on transmet ensuite à un opérateur central qui élabore la décision.

Les systèmes que nous étudions ont pour mission de résoudre un problème de détection c'est-à-dire un problème de choix entre deux hypothèses H_0 (bruit seul) et H_1 (signal plus bruit) à partir d'observations collectées sur un ensemble de N sources d'information. Chaque source délivre un vecteur x_i , $i = 1, \dots, N$ et l'information globale X est la concaténation de tous ces vecteurs. Selon les cas les vecteurs x_i peuvent être des échantillons pris à la sortie d'un capteur à des instants successifs, les mesures d'un phénomène observé à l'aide de capteurs de nature différente (radar, rayon X, infra-rouge ...) ou encore les observations de capteurs implantés sur différents sites géo-



graphiques. Chaque vecteur \mathbf{X} est associé à un point de l'espace E produit des espaces d'observation des capteurs $E = \prod E_i$. Chaque vecteur \mathbf{X} est également une réalisation particulière de la variable aléatoire X dont la loi de probabilité dépend de l'hypothèse. Les hypothèses sont dites simples lorsque les lois de probabilité conditionnelles de l'observation $P_0(\mathbf{X}) = P(\mathbf{X} | H_0)$ et $P_1(\mathbf{X}) = P(\mathbf{X} | H_1)$ sont connues.

2. APPROCHES PROBABILISTES

Ce type d'approche correspond à l'optimisation des systèmes au sens d'un critère fondé sur la description probabiliste des phénomènes. Les systèmes comprennent des processeurs locaux notés γ_i qui élaborent chacun à partir d'une observation x_i un résumé local $u_i = \gamma_i(x_i)$ et l'opérateur de fusion F qui construit la sortie du système en combinant les résumés locaux $F(u_1, \dots, u_n)$. L'ensemble regroupant la loi de fusion et les règles de traitement locales $\Gamma = \{\gamma_1, \dots, \gamma_n, F\}$ constitue la stratégie du système. Pour résoudre le problème de la détection décentralisée nous devons optimiser la stratégie Γ , c'est-à-dire optimiser un ensemble de fonctions. Deux classes de solutions sont possibles selon que les résumés sont continus ou qu'ils prennent leur valeur dans un alphabet dénombrable.

A. Système centralisé optimal

La solution à structure libre proposée par la théorie classique de la détection calcule le rapport de vraisemblance global de l'observation $\Lambda(\mathbf{X})$ défini comme le rapport des probabilités conditionnelles et le compare à un seuil η .

$$\Lambda(\mathbf{X}) = \frac{P_1(\mathbf{X})}{P_0(\mathbf{X})} \underset{<}{\overset{>}{\geq}} \eta. \quad (1)$$

Lorsque les observations sont indépendantes sous H_0 et sous H_1 , le rapport de vraisemblance global se factorise simplement comme le produit des rapports de vraisemblance locaux. Dans ce cas particulier le système optimum à structure libre peut être réalisé de façon décentralisée en calculant sur chaque site les rapports de vraisemblance locaux $\gamma_i(x_i) = \Lambda(x_i)$ et en les transmettant intégralement au processeur central qui effectue le produit et le seuillage.

B. Quantification répartie

La quantification répartie est caractérisée par le fait qu'un quantificateur vectoriel est associé à chaque capteur. Elle appartient à la classe des systèmes décentralisés dont les résumés locaux sont à valeurs discrètes. Les études sur l'optimisation de cette structure [1], [2] ont établi que lorsque les observations sont indépendantes sous les hypothèses H_0 et H_1 la structure optimale de quantification locale est la quantifi-

cation par rapport de vraisemblance pour laquelle on quantifie non pas l'observation mais son rapport de vraisemblance calculé au préalable. Malheureusement, en l'absence de toute hypothèse d'indépendance des observations, l'optimisation globale des systèmes de quantification répartie est un problème très complexe qui ne possède pas de solution explicite.

C. Détection décentralisée avec réduction de dimension

A l'inverse de l'approche précédente, la détection décentralisée avec réduction de dimension ne transforme pas les vecteurs observation en grandeurs discrètes mais en grandeurs scalaires, les processeurs locaux et l'opérateur de fusion sont donc des filtres. Cette approche repose sur l'hypothèse selon laquelle les vecteurs observation recueillis par les capteurs sont de dimension bien supérieure à un. Dans ce cas on peut considérer qu'une économie importante en termes de coût de transmission est réalisée dans la mesure où les résumés émis par les processeurs locaux sont des quantités scalaires. A l'exception du cas particulier de l'indépendance des observations, où comme on l'a vu au paragraphe A, les filtres locaux doivent calculer les rapports de vraisemblance $\Lambda(x_i)$ que la fusion multiplie simplement, le problème reste trop général et extrêmement difficile à résoudre. On est donc amené à introduire une simplification qui trouve son origine dans la constatation que le rapport de vraisemblance global peut être approximé par un filtre plus structuré comme par exemple les filtres de Volterra qui s'écrivent comme une somme de termes. C'est pour cette raison que la fusion envisagée pour cette approche est une simple addition [3]. On démontre alors que le processeur local optimal est de la forme

$$\gamma_i(x_i) = \Lambda_c(x_i) + \text{Fonc}\{P_0(\mathbf{X}), \gamma_j(x_j), j \neq i\}, \quad (2)$$

où $\Lambda_c(x_i) = \Lambda(x_i) - 1$ est le rapport de vraisemblance déplacé auquel s'ajoute un second membre qui traduit l'influence des autres processeurs locaux et des propriétés statistiques de l'observation.

3. L'APPROCHE NEURONALE

La synthèse des différentes structures basées sur des traitements conventionnels probabilistes nous a permis de mettre à jour quelques enseignements que nous allons exploiter pour concevoir l'approche neuronale. Tout d'abord le rôle particulier que joue le rapport de vraisemblance présent dans toutes les solutions des approches probabilistes. Il est apparu également que l'optimisation des structures conventionnelles n'est accessible qu'au travers d'hypothèses simplificatrices telles la linéarité de la règle de fusion ou l'indépendance des observations. Or cette dernière hypothèse n'est réaliste que lorsque les bruits sont indépendants d'un

capteur à l'autre et que le signal à détecter est connu. Dès que l'une ou l'autre de ces conditions est fautive, l'hypothèse n'est plus valable et seuls des systèmes sous-optimaux peuvent être développés. L'objectif de l'approche neuronale est d'offrir un cadre dans lequel il soit possible de développer des systèmes de détection décentralisée s'affranchissant des hypothèses simplificatrices habituelles.

De façon générale l'objectif d'un système de détection décentralisée est de remplacer le plus efficacement possible le rapport de vraisemblance global $\Lambda(X)$ par une structure décentralisée moins coûteuse en termes d'acheminement des données. Cette démarche peut s'envisager en termes d'apprentissage du fonctionnement du système centralisé optimum par une structure décentralisée. Lorsque l'on parle d'apprentissage il est naturel d'aborder le sujet des réseaux de neurones dont le fonctionnement est fondé sur ce principe.

Les conditions d'utilisation des réseaux neuronaux ne répondent pas à un formalisme bien défini et leur utilisation est bien souvent fondée sur une heuristique. La référence [4] fixe un cadre général en proposant une liste de caractéristiques d'application. Une première caractéristique veut que le problème fasse intervenir des données bruitées, ce qui est intrinsèque au problème de la détection. Une autre caractéristique importante consiste à vérifier que la solution du problème est inconnue ou difficile à expliciter mais que l'on dispose des solutions qui correspondent à un ensemble d'entrées tests. Cette caractéristique est bien vérifiée par le problème de la détection décentralisée puisque l'on connaît la solution exacte qui n'est autre que le rapport de vraisemblance centralisé qui constitue un signal d'apprentissage idéal. De plus à chaque valeur du rapport de vraisemblance centralisé correspondent un ou plusieurs vecteurs dont les composantes sont les rapports de vraisemblance locaux. Il existe donc une relation non triviale entre ces quantités. Ne disposant pas de l'expression de cette relation nous allons tenter de l'apprendre à un neurone de type adaline dont les entrées sont des fonctions non linéaires des rapports de vraisemblance locaux.

Le choix d'un seul neurone de type adaline présente deux avantages. Il nous permet tout d'abord de montrer de façon très simple l'intérêt d'introduire des méthodes connexionnistes dans les systèmes de détection décentralisée. De plus, l'une des propriétés du neurone adaline est qu'il est capable de reconnaître une frontière non linéaire entre deux sous-espaces par combinaison de termes non linéaires de l'entrée. Or l'interprétation géométrique du problème de la détection est la division de l'espace d'observation global E en deux domaines de décision D_0 et D_1 tels que si l'observation tombe dans D_0 (respectivement D_1) la décision est H_0 (respectivement H_1). L'apprentissage de la fonction "rapport de vraisemblance" est donc équivalent à l'apprentissage d'une surface de séparation entre les domaines D_0 et D_1 . Nous allons voir sur un exemple que cet apprentissage est possible.

4. EXEMPLE

Un des cas les plus courants pour tester ce genre de dispositif consiste à détecter un signal d'amplitude constante S dans un bruit gaussien normé, centré. Nous allons simuler une structure de détection décentralisée avec apprentissage comportant deux capteurs recevant chacun un vecteur x_1 et x_2 de dimension trois. Notons que dans le cas où les observations sont indépendantes d'un capteur à l'autre, le rapport de vraisemblance déplacé centralisé s'écrit :

$$\Lambda_c(X) = \Lambda_c(x_1) + \Lambda_c(x_2) + \Lambda_c(x_1)\Lambda_c(x_2). \quad (3)$$

Dans une démarche qui vise à montrer l'intérêt d'introduire des méthodes d'apprentissage dans le problème de la détection décentralisée, la première étape est tout naturellement de montrer qu'il est possible d'apprendre la relation (3) pour des observations indépendantes. Nous référant à l'importance du rapport de vraisemblance dans les problèmes de détection décentralisée, le système que nous étudions calcule localement le rapport de vraisemblance déplacé de chaque capteur du réseau, $\gamma_i(x_i) = \Lambda_c(x_i)$. Ces résultats intermédiaires sont transmis au niveau central où ils sont associés par la fonction produit. Un neurone combine ensuite les trois termes $\Lambda_c(x_1)$, $\Lambda_c(x_2)$ et $\Lambda_c(x_1)\Lambda_c(x_2)$ dans une somme pondérée $\omega_1\Lambda_c(x_1) + \omega_2\Lambda_c(x_1)\Lambda_c(x_2) + \omega_3\Lambda_c(x_2)$, pour élaborer la décision d'ensemble. La méthodologie de cette solution est de dissocier la partie "réduction de l'information" qui est mise en oeuvre par un traitement conventionnel probabiliste de la phase de "prise de décision" qui met en oeuvre un neurone.

La configuration du système est représentée sur la figure 1. L'apprentissage se déroule de la façon suivante : on présente au neurone des rapports de vraisemblance locaux correspondant alternativement à des situations avec ou sans signal, la sortie du neurone est comparée au rapport de vraisemblance centralisé calculé par ailleurs pour obtenir le signal d'erreur utilisé par l'algorithme LMS d'adaptation des poids du neurone. L'apprentissage est stoppé après une succession de cinq séquences de 1000 échantillons correspondant à la situation bruit seul suivis de 1000 échantillons correspondant à la présence du signal. A l'issue de l'apprentissage les poids du neurone sont fixés et l'on teste la capacité de généralisation du système en lui faisant "suivre" l'apparition et la disparition d'un signal constant d'amplitude 0,1 noyé dans le bruit gaussien centré normé. A l'issue de la présentation des 2000 échantillons du signal d'apprentissage les poids convergent vers les valeurs $\omega_1 = 0.9474$, $\omega_2 = 0.9051$ et $\omega_3 = 0.9583$. Bien entendu lors des phases de fonctionnement le rapport de vraisemblance centralisé n'est pas calculé et l'entrée correspondante est déconnectée.

La figure 2 représente des résultats obtenus en phase de fonctionnement avec les poids donnés ci-dessus. Les figures



2.a, 2.b, 2.c et 2.d représentent respectivement le signal à détecter, le rapport de vraisemblance centralisé $\Lambda_c(X)$, l'estimation de $\Lambda_c(X)$ obtenue à la sortie du neurone, l'erreur commise sur l'estimation de $\Lambda_c(X)$ et enfin le signal détecté par notre système. Les performances du système sont représentées sur la figure 3 à l'aide d'une courbe C.O.R. qui compare les taux de détection P_d et de fausse alarme P_{fa} obtenus dans les circonstances de l'exemple par le système centralisé (courbe continue) et notre système décrit sur la figure 1. (repérés par le marqueur +).

Enfin 50 apprentissages sur 2000 échantillons ont été effectués, qui nous ont permis de faire un calcul de moyenne m_i et d'écart-type σ_i sur les valeurs que prennent les poids à la fin de chaque apprentissage. Les résultats obtenus sont : pour le poids ω_1 $m_1 = 0.9908$ et $\sigma_1 = 5.75 \cdot 10^{-2}$, pour le poids ω_2 $m_2 = 0.9816$ et $\sigma_2 = 1.01 \cdot 10^{-1}$ et enfin pour le poids ω_3 $m_3 = 0.9880$ et $\sigma_3 = 5.08 \cdot 10^{-2}$.

5. CONCLUSION ET PERSPECTIVES

Les résultats d'une étude plus générale utilisant un réseau de neurones seront bientôt disponibles. L'intérêt des réseaux de neurones est qu'ils peuvent approximer n'importe quelle fonction mesurable en utilisant simplement des exemples de cette fonction. En effet il a été démontré [5] qu'un réseau multi-couches à une couche cachée de neurones sigmoïdes est capable d'approximer toute fonction mesurable. Nous allons bien entendu exploiter cette propriété pour l'approximation du rapport de vraisemblance global.

Références

- [1] M.P. Boyer and B. Picinbono, "Quantization and distributed detection", Proc ICASSP, pp. 2688-2691, May 1989.
- [2] M.P. Boyer, "Détection et estimation réparties, approches par la quantification et le filtrage de Volterra", Thèse de doctorat, Université de Paris Sud, 1992.
- [3] B. Picinbono and M.P. Boyer, "A new approach of decentralized detection", Proc ICASSP, May 1991.
- [4] E. Davalo, P. Naim, "Des reseaux de neurones", Eyrolles, 1989.
- [5] K. Hornik, M. Stinchcombe, H. White, "Multilayer feed-forward networks are universal approximators", *Neural Networks*, vol 2, no5, 359-366, 1989.

Remerciements : Cette étude a été menée au sein du groupe ERASME soutenu par de la Fédération Universitaire et Polytechnique de Lille, le FEDER et le conseil régional Nord-Pas de calais.

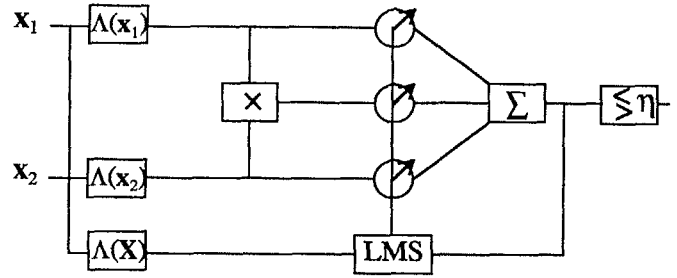


Figure 1

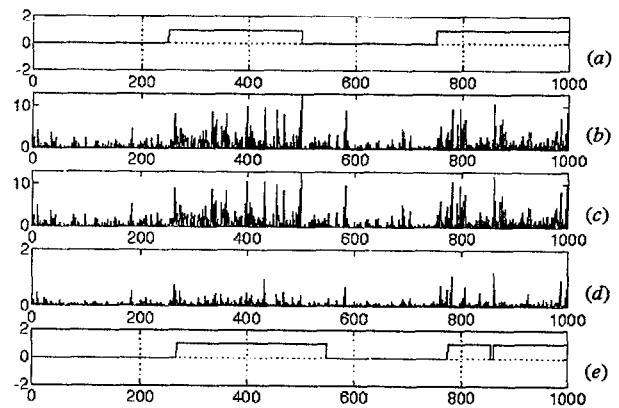


Figure 2

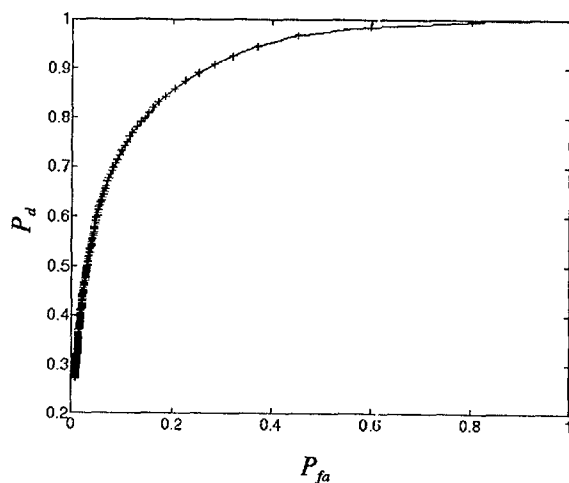


Figure 3