



CLASSIFICATION D'HELICOPTERE EN VOL PAR RESEAU DE NEURONES

B.BALIGAND

METRAVIB R.D.S. 64, Chemin des Mouilles
B.P 182 - 69132 ECULLY Cédex

RÉSUMÉ

On aborde le problème de la classification automatique de signature acoustique par réseaux de neurones (R.N.). L'opération consiste à reconnaître une classe d'événements C_i parmi K classes possibles à partir de paramètres descripteurs \underline{x} . L'article analyse trois architectures de R.N. fondamentalement différentes et les compare à l'approche bayésienne. On étudie deux classifieurs supervisés (le MBPN et le CPN) de type réseau approximateur et un classifieur non supervisé (l'ART) de type réseau auto-organisant. Des essais sont d'abord réalisés sur des distributions multi-normales, puis sur des données réelles provenant d'hélicoptères en vol. Les taux de réussite obtenus en reconnaissance d'hélicoptère (> 90 %) illustrent le potentiel des R.N. eu égard à la nature multipolaire des sources et aux configurations de vol.

I. INTRODUCTION

Le problème traité est celui de la **classification par réseaux de neurones (R.N.) pour la reconnaissance d'hélicoptère en vol**. Le problème générique posé par la classification est de trouver une application \emptyset qui, à tout vecteur \underline{x} ($\underline{x} \in \mathbb{R}^N$) associe une région de décision $\emptyset(\underline{x}) = 1, 2, \dots, K$ pour K régions possibles.

Dans le contexte bayésien, l'application \emptyset est calculée à partir des fonctions densité de probabilité conditionnelle de \underline{x} et le classifieur est optimal. Dans le contexte neuronal, les hypothèses sur la statistique de \underline{x} ne sont plus nécessaires et la classification est obtenue après convergence vers une estimée de \emptyset .

L'article est essentiellement organisé en deux parties. La première est consacrée à l'analyse de trois architectures de R.N. (MBPN, CPN, ART) et la seconde est consacrée à leurs applications sur des données simulées, puis sur des données acoustiques mesurées à partir d'hélicoptères en vol.

II. CLASSIFICATION BAYESIENNE

Le classifieur bayésien est le premier outil décisionnel utilisé. Il repose sur une stratégie probabiliste optimale et demeure un outil de base qui, sur le concept, sert de préambule à l'approche neuronale. La classification d'un vecteur \underline{x} ($\underline{x} \in \mathbb{R}^N$) parmi K classes possibles (C_1, C_2, \dots, C_K) revient à déterminer les fonctions séparatrices $\theta^N(\underline{x})$ des K domaines \mathcal{D}_k de \mathbb{R}^N . Dans le contexte

ABSTRACT

This article deals with the use of Neural Network (N.N.) in classification of acoustical signals. It is a pattern classification problem of the realisations class C_i among K possible classes with the parameters vector \underline{x} of the realisation. We analyse three Neural Network architectures that are fundamentally different and they are compared to the bayesian classifier. We study two supervised learning N.N. (the MBPN and the CPN) and one non supervised learning N.N. (the ART). These classifiers are applied on multi-normal distributions and then applied on real data coming from flying helicopters. The good helicopter recognition results (> 90 %) show the potentiality of the N.N.

bayésien, ces fonctions séparatrices de \mathbb{P}^N sont estimées en minimisant une fonction risque de nature probabiliste, la stratégie étant de pénaliser une mauvaise classification. On définit ainsi une fonction coût qui, à tout couple d'éventualité (état vrai, état choisi), (i, j) , de $\{1, 2, \dots, K\}^2$, associe un réel $c(i, j)$ tel que $c(i, j) > 0$ si $i \neq j$ et $c(i, j) = 0$ si $i = j$. Une fonction risque R est ensuite définie comme l'espérance mathématique du coût sur l'ensemble des observations toutes classes confondues. On a ainsi :

$$(2-1) \quad R = \sum_{i \neq j} c(i, j) \cdot P(j) \cdot \int_{\mathcal{D}_i} p(\underline{x}/C_i) \cdot d\underline{x}$$

où : $P(j)$ est la probabilité a priori qu'une observation \underline{x} appartienne à la classe C_j

$p(\underline{x}/C_i)$ est la Fonction Densité de Probabilité conditionnelle de \underline{x} (F.D.P.).

La minimisation de la fonctionnelle R conduit aux séparatrices $\theta^N(\underline{x})$ définissant ainsi au sens du critère de Bayes un classifieur optimal.

Pour illustrer ceci, on considère le cas de deux classes ($K = 2$) à distribution gaussienne de même covariance $\underline{\Gamma}_1 = \underline{\Gamma}_2 = \underline{\Gamma}$.

$$(2-2) \quad p(\underline{x}/C_k) = \frac{1}{(2\pi)^N \det[\underline{\Gamma}]} \cdot e^{-\frac{1}{2} (\underline{x} - \underline{m}_k)^t \cdot \underline{\Gamma}^{-1} \cdot (\underline{x} - \underline{m}_k)}$$

$k = 1$ ou 2 .



où : \underline{m}_1 et \underline{m}_2 sont les centres de $p(\underline{x}/C_1)$ et $p(\underline{x}/C_2)$
 $\det[\underline{\Gamma}]$ est le déterminant de $\underline{\Gamma}$

avec équiprobabilité a priori sur les deux classes C_1 et C_2 des calculs simples [1] fournissent comme meilleure séparatrice dans \mathbb{R}^N l'hyperplan :

$$(2-3) \quad \theta^N(\underline{x}) = \underline{\omega}^t \cdot \underline{x} + \omega_0 = 0 \text{ avec } \begin{cases} \underline{\omega} = \underline{\Gamma}^{-1} \cdot (\underline{m}_2 - \underline{m}_1) \\ \omega_0 = \frac{1}{2} (\underline{m}_1 - \underline{m}_2)^t \cdot \underline{\Gamma}^{-1} \cdot (\underline{m}_1 + \underline{m}_2) \end{cases}$$

Ces résultats sont directement liés aux deux premiers moments statistiques des FDP $p(\underline{x}/C_i)$. En pratique, ces moments sont très souvent estimés au sens du maximum de vraisemblance à partir d'un ensemble d'apprentissage contenant p réalisations : $\mathcal{B}_p = \{ [\underline{x}^\ell, \underline{c}(\underline{x}^\ell)] \}$, $\ell = 1, \dots, p$. A la phase d'apprentissage succède une phase d'évaluation à partir d'un ensemble de généralisation contenant q réalisations : $\mathcal{B}_q = \{ [\underline{x}^\ell, \underline{c}(\underline{x}^\ell)] \}$, $\ell = 1, \dots, q$.

III. CLASSIFICATION PAR RÉSEAUX DE NEURONES

Dans le contexte neuronal, les connaissances et hypothèses sur les FDP $p(\underline{x}/C_i)$ ne sont plus nécessaires. De même, la classe des fonctions séparatrices $\theta^N(\underline{x})$ n'est plus limitée au seul cas linéaire (ou quadratique) et peut être étendue à des formes plus complexes. La classification par R.N. est alors réalisée par une approximation de \mathbb{R}^N dans \mathbb{R} après convergence vers une estimée de $\theta^N(\underline{x})$.

III.1. Le réseau M.B.P.N. ("Multilayer Backpropagation")

C'est un **modèle multicouche** entièrement connecté fonctionnant par **apprentissage supervisé**. La figure 1 en donne la structure à une couche cachée.

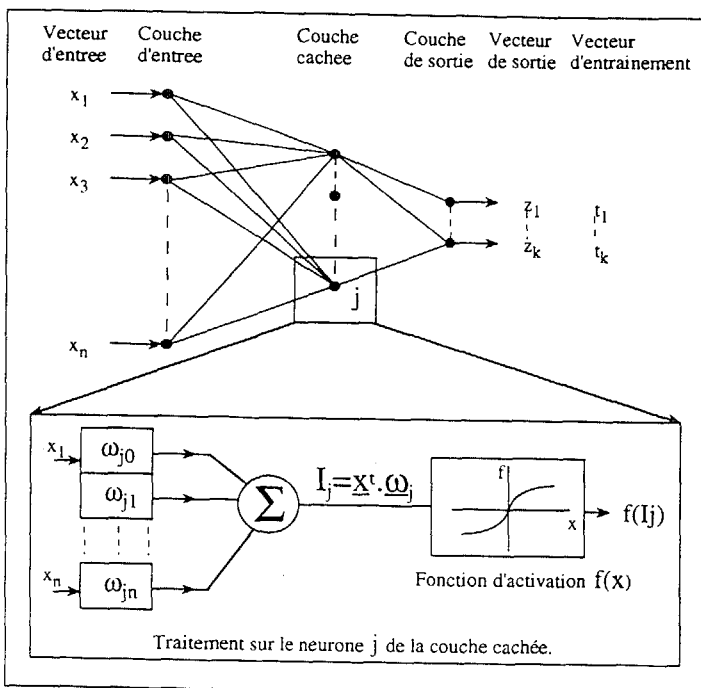


Figure 1 : Configuration du multicouche type MBPN

La règle d'apprentissage est un algorithme de **rétropropagation du gradient de l'erreur**. Les poids $\underline{\omega}$ des connexions sont modifiés selon une règle du gradient due à WIDROW [2]. Ainsi, pour chaque neurone i à j connexions, on a pour la dernière couche :

$$(3-1) \quad d\omega_{ij}/dt = \alpha \cdot \delta_i \cdot z_j$$

avec :

$$\begin{cases} \delta_i = f'(I_i) \cdot (t_i - z_i) & (f'(x) = \partial f / \partial x) \\ z_j = f(I_j) \\ \alpha = \text{coefficient de convergence} \end{cases}$$

Propriété des R.N. multicouches

- On montre [5] qu'un R.N. multicouches peut toujours approcher au sens de WEIERSTRASS une fonction continue quelconque ($\psi : \mathbb{R} \rightarrow \mathbb{R}$) sur le segment $[a, b]$ avec une précision définie par le théorème de BOREL. De même, ce type de R.N. permet d'approximer une relation quelconque ($\emptyset : \mathbb{R}^n \rightarrow \mathbb{R}^k$) avec une précision arbitrairement grande. Cela suppose que le R.N. possède au moins une couche cachée avec suffisamment de neurones à fonction d'activation sigmoïde $f(x) = (1 + e^{-b \cdot x})^{-1}$.

III.2. Le réseau C.P.N. ("Counterpropagation") [3]

C'est un R.N. de type approximateur fonctionnant par **apprentissage supervisé** avec une règle qui est basée sur un **critère de distance et une compétition**. Sa structure est constituée d'une couche compétitive de KOHONEN et d'une couche approximante de GROSSBERG.

- La couche de KOHONEN constitue un R.N. auto-organisant où l'ajustement des poids des connexions ω se fait par compétition. Le résultat fournit les k neurones plus proches voisins de \underline{x} parmi les N vecteurs $\underline{\omega}_j$ mémorisés. La compétition est régie par une distance euclidienne d_j .

$$(3-2) \quad d_i = \|\underline{\omega}_i - \underline{x}\| = \sqrt{\sum_{j=1}^n (\omega_{ij} - x_j)^2}$$

Les poids des connexions sont alors ajustés par une règle de type gradient :

$$(3-3) \quad \begin{cases} d\omega_j / dt = \alpha \cdot (\underline{x} - \underline{\omega}_j) & \text{si } \omega_j \in k \text{ plus proches voisins} \\ d\omega_j / dt = 0 & \text{si non} \end{cases}$$

où α est le coefficient de convergence de la couche de KOHONEN. La couche de GROSSBERG affecte l'entrée \underline{x} à la décision finale z_i à partir des sorties de la couche de KOHONEN et des poids u_{ij} ajustés à l'apprentissage selon une règle du type gradient.

$$(3-4) \quad \begin{cases} z_i = \sum_{j=1}^k u_{ij} \\ du_{ij} / dt = a (t_i - u_{ij}) \end{cases}$$

avec : t_i : sortie réelle ($1 \leq i \leq k$) contenue dans les neurones d'apprentissage
 a : coefficient de convergence de la couche de GROSSBERG

Propriétés du réseau C.P.N.

- La fonction de densité des points des vecteurs ω_i de la couche de KOHONEN tend vers la densité de probabilité $p(\underline{x}/\omega_k)$ de l'entrée \underline{x} et les ω_i s'ordonnent de manière à conserver les relations topologiques présentes sur les vecteurs d'entrée. Le réseau CPN constitue, de ce fait, un estimateur de Fonction Densité de Probabilité [3].
- La convergence du réseau C.P.N. est plus rapide que celle du MBPN.

III.3. Le réseau ART ("Adaptive Resonance Theory") [4]

Ce R.N. se situe dans le contexte des réseaux à **apprentissage non supervisé**. Il se démarque des deux R.N. précédents en satisfaisant aux contraintes suivantes :

- être adaptatif et cependant autostabilisant, donc apte à résoudre le dilemme plasticité/stabilité : **l'apprentissage et la généralisation sont exécutés simultanément**,
- être auto-organisant : le nombre de classes n'est pas fixé au départ,
- être capable de créer de nouvelles classes.

La structure étudiée est celle proposée par CARPENTER et GROSSBERG [4]. L'ART se compose de quatre parties (voir figure 2) :

- le champ des caractéristiques (couche F1) : il mémorise les vecteurs d'entrée \underline{x}
- le champ des catégories (couche F2) : il représente les "clusters" de sortie,
- le poids des interconnexions entre les deux couches : u_{ij} pour le filtre de bas en haut et v_{ij} pour le filtre de haut en bas. Ils mémorisent les informations statistiques après apprentissage.
- le système de surveillance qui contrôle la convergence.

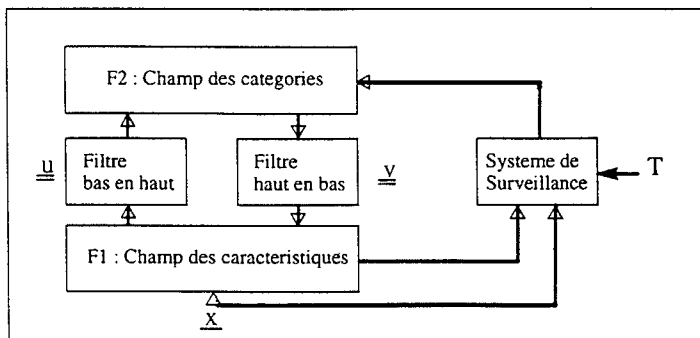


Figure 2 : Configuration générale de l'ART

Le fonctionnement de l'ART se fait en quatre étapes :

- 1) **Excitation** : Un vecteur \underline{x} se présente sur F1, est comparé aux prototypes établis précédemment u_{ij} par critère de corrélation

$$(3-5) \quad X_j = \underline{x}^t \cdot \underline{u}_j$$

- 2) **Compétition** : Les neurones de F2 entrent en compétition pour déterminer la catégorie la plus proche de \underline{x} au sens du maximum de corrélation

$$(3-6) \quad Y_j \begin{cases} = 1 & \text{si } X_j = \max [X_i, i = 1, 2, \dots, M] \quad M \text{ prototypes} \\ = 0 & \text{si non} \end{cases}$$

- 3) **Réponse** : Le neurone élu j de F2 agit à son tour sur ceux de F1 par les connexions v_{ij} . Les neurones de F1 vont alors converger vers la sortie \underline{X}

$$(3-7) \quad \underline{X} = \underline{x} \cap \underline{v}_j \quad \text{où } \cap \text{ est l'opération d'intercorrélation}$$

- 4) **Contrôle** : Le système de surveillance arbitre la convergence en mesurant la ressemblance entre \underline{x} et \underline{v}_j par intercorrélation normalisée

$$(3-8) \quad t = \|\underline{x} \cap \underline{v}_j\| / \|\underline{x}\| \quad \text{et } T \text{ seuil admissible}$$

- si $t < T$, aucun prototype établi n'est satisfaisant pour représenter \underline{x} , le neurone j de F2 est alors neutralisé et un nouveau prototype est établi,
- si $t > T$, le prototype sélectionné est considéré comme représentant de \underline{x} , le réseau entre dans un état résonnant et modifie les poids associés avec une règle qui est encore du type gradient.

IV. ANALYSE COMPARATIVE DES CLASSIFIEURS SUR DES DONNÉES

Pour évaluer les performances de ces R.N. utilisés en classifieurs, nous avons réalisé des essais comparatifs d'abord sur des distributions multi-normales puis sur signatures acoustiques provenant d'hélicoptères en vol. Tous les résultats sont exprimés en **taux de bonne classification** avec des configurations de R.N. dont les paramètres ont été optimisés.

IV.1. Données simulées à distribution gaussienne

On compare les performances du MBPN et de l'ART sur des variables aléatoires gaussiennes pour des cas linéairement séparables. On présente les résultats de 2 essais dichotomiques ($k = 2$) ; le premier est de dimension 2 ($N = 2$) et le second est de dimension 6 ($N = 6$).

Pour chacun des essais :

- on donne le résultat optimal obtenu par le classifieur bayésien,
- les R.N. sont entraînés sur 400 vecteurs tirés aléatoirement et la généralisation est faite sur 10000 vecteurs.

essai 1 : $\underline{m}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ $\underline{m}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ $\underline{\Gamma}_1 = \underline{\Gamma}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

essai 2 : $\underline{m}_1^t = (1,0, 1,0, 1,0)$ $\underline{m}_2^t = (0,1, 0,1, 0,1)$ $\underline{\Gamma}_1 = \underline{\Gamma}_2 = \text{diag}(1)$

MBPN1 : entrée = 2 neurones sortie = 1 neurone $f(x) = \text{sigmoïde}$

MBPN2 : entrée = 2 neurones couche cachée = 6 neurones

sortie = 1 neurone $f(x) = \text{sigmoïde}$



Pour évaluer l'ART on attribue chaque catégorie à la classe majoritairement représentée (> 50 %) par les vecteurs dont elle est constituée.

On obtient les taux de réussite suivants :

	MBPN1	MBPN2	ART	Clas. Bayésien
Essai 1	66,9 %	54,6 %	72,5 %	76 %
Essai 2	86,6 %	62,7 %	84,0 %	89 %

La meilleure configuration MBPN est celle à une seule couche. Cela s'explique dans ce cas par la séparatrice $\theta^2(x)$ qui est une droite du type $\omega_0 + \omega_1 \cdot x_1 + \omega_2 \cdot x_2 = 0$. La convergence de l'ART est beaucoup plus rapide que celle du MBPN. L'ART fournit à 5 % près le résultat optimal au prix d'un nombre de sous-classes élevé (≥ 6).

IV.2. Données réelles de signatures acoustiques d'hélicoptères en vol

Cette seconde application concerne l'identification d'hélicoptères en vol par R.N. [6] à partir de leur signature acoustique. L'intérêt de ce problème tient à la **grande variabilité des signatures qui sont fonction de la position de l'hélicoptère** et nous situe dans un cas **non linéairement séparable**. L'analyse des signaux nous a conduit à extraire 26 paramètres jugés discriminants. Ils sont issus de trois types de description : temps/échelle (harmoniques du rotor principal et du rotor de queue), temporelle (facteurs de forme et coefficient cepstral), fréquentielle (puissance en bande tiers d'octave).

On compare les performances du MBPN et du CPN sur ces 26 paramètres pour différentes configurations de vol. La règle générale adoptée est de réaliser **un apprentissage en situation de champ proche (à 300 m)** avec un bon rapport signal à bruit, puis **une généralisation en champ lointain** à plus de 2 km.

On obtient les taux de réussite suivants :

Configurations à la généralisation		MBPN	CPN
Hélicoptère	C_1 position A à 300 m	100 %	100 %
	C_1 position B à plus de 2 km	96 %	100 %
Hélicoptère	C_2 position C à plus de 2 km	100 %	100 %
	C_2 position D à plus de 2 km	100 %	100 %
Hélicoptère	C_3 position E à plus de 2 km	100 %	100 %
	C_3 position F à plus de 2 km	98 %	100 %
Hélicoptère	C_4 position G à plus de 2 km	98 %	100 %
	C_4 position H à plus de 2 km	98 %	96 %
Vent et pluie	C_5	83 %	90 %

On présente les résultats pour 4 classes d'hélicoptères (C_1 , C_2 , C_3 , C_4) et une classe C_5 correspondant à du vent ou de la pluie.

L'apprentissage est fait sur 5 réalisations par classe et la généralisation porte sur 48 réalisations par classe.

Le réseau CPN qui converge plus vite, fournit en moyenne (à moins de 5 % près) un taux de réussite voisin du MBPN. Comme le montre le tableau ci-dessus, la reconnaissance des quatre hélicoptères et du bruit par les 2 R.N. est très bonne. Ceci témoigne de l'aptitude d'une telle approche à traiter les deux aspects suivants :

- invariance vis-à-vis de l'orientation de l'hélicoptère par rapport à l'antenne,
- capacité à la généralisation en champ lointain à partir d'un apprentissage en champ proche.

VI. CONCLUSION

L'analyse comparative de 3 R.N. (MBPN, CPN, ART) sur une base de données simulées et réelles révèle les limites de ces outils utilisés en classifieurs (convergence lente, solutions multiples, choix des paramètres) mais permet également de souligner leurs atouts (pas d'hypothèse statistiques, pas d'inversion matricielle, caractère adaptatif, solutions non linéaires).

On remarquera la structure particulière du réseau ART et les principes qui président à sa construction. A ce titre, il semble receler de grandes potentialités (autonomie, apprentissage rapide) mais demande toutefois à être supervisé pour l'affectation de sorties.

REMERCIEMENTS

L'auteur remercie J.M. PAROT (IMDYS) pour ses conseils et suggestions ainsi que L. PIN, stagiaire DEA ICPI 92, pour les simulations sur les distributions gaussiennes.

RÉFÉRENCES

- [1] R.O. DUDA, P.E. HART - "Pattern classification and scene analysis" - WILEY 1973.
- [2] B. WIDROW and M.A. LEHR - "30 years of adaptive Neural Networks : Perceptron, Madaline and Backpropagation" - IEEE, vol 78, n°9, September 1990.
- [3] R. HECHT-NIELSEN - "Counterpropagation Networks" - Applied Optics, pp 4979-4984, 26, (1987).
- [4] G.A. CARPENTER, S. GROSSBERG - "A massively parallel architecture for a self-organizing neural pattern recognition machine" - Computer Vision, 37, 54-115 (1987).
- [5] ZHANG WEI, YU YINGLIN - "Approximation property of multi-layer network (MLNN) and its application in nonlinear simulation" - pp 171-176, IJCNN, Seattle 1991.
- [6] B. BALIGAND, J. LARCHER - "Outils décisionnels pour la surveillance acoustique et vibratoire" - pp 555-562, Conf. Int. CETIM, SFM/SFA, 1992.