Equipe : Réseaux de Neurones

Composition de l'équipe : (99)

F. Badran (Maître de conférences CNAM),

Visiteur S.Thiria (Professeur université de Versailles St Quentin en Yvelines),

Nabil Metoui (Doctorant).

Deux étudiants ingénieurs : B. MABBOUX et D. FRAYSSINET.

Composition de l'équipe : (2000)

F. Badran (Maître de conférences CNAM).

Visiteur S.Thiria (Professeur université de Versailles St Quentin en Yvelines),

Responsable : Fouad Badran
Tél. : 01 40 27 22 69
Télécopie: 01 40 27 27 09
E_Mail : badran@cnam.fr
thiria@cnam.fr

Description des activités de recherche

Objectifs

Nos activités de recherches sont centrées sur l'étude et la mise en pratique des méthodes neuronales. Nos travaux sont organisés selon deux axes complémentaires : études théoriques et traitement d'applications réelles.

La partie théorique porte principalement sur :

- La modélisation des fonctions de transfert ainsi que du bruit de mesures à partir de données réelles.
- Le filtrage du bruit et l'élimination des points aberrants lors de l'apprentissage d'un réseau multicouche.
- La résolution de problèmes inverses mal posés qui intègrent les méthodes adjointes et la prise en compte des corrélations spatiales.
- L'assimilation des mesures observées dans les modèles numériques (notamment en océanographie).
- L'utilisation de la version probabiliste des cartes auto-organisatrices (PRSOM) pour la classification notamment lorsqu'on ne dispose que peu de données étiquetées par un expert.
- L'étiquetage des cartes topologiques par des méthodes de classification hiérarchiques. Développement d'une famille de mesures de similarités permettant de prendre en compte la distance euclidienne dans l'espace des données et la conservation de la topologie dans l'espace discret de la carte.
- Développement d'un algorithme de type Carte Topologique adapté aux données binaires et adaptation du modèle de mélange des lois de Bernoulli pour le calcul des probabilités.

En ce qui concerne les applications, nous avons abordé les problèmes suivants :

- Calcul du vent marin par inversion des données des diffusiomètres satellitaires ERS1/2 et NSCAT (avec le Laboratoire d'Océanie DYnamique et de Climatologie LODYC).
- Classification de nuages à partir d'images Météosat (avec le LODYC et le LMD)
- Inversion des mesures radiométriques sur la couleur de l'océan (avec le LODYC).
- Détermination de cartes isobathes représentant la topographie du fond de l'océan,

application à la Méditerranée (avec le LODYC).

- Assimilation des données dans les modèles Océanique de biochimie (avec le LODYC).
- Traitement des mesures de diagraphies en vue de déterminer l'épaisseur et la nature des couches géologiques traversées par le forage ainsi que pour retrouver certaines caractéristiques non mesurées (porosité perméabilité) (avec ELF et le laboratoire de Géochronologie, Géochimie et Pétrologie de la croûte et de ses interfaces de l'Institut des Sciences de la Terre, de l'Eau et de L'Espace de Montpellier)
- Classification automatique des molécules (Avec L'OREAL et le groupe réseau de neurone de LIPN de l'université de Paris 13).

Résultats obtenus

1 . Modélisation de fonctions de transfert

Nous avons défini une méthode qui permet de calculer la fonction du modèle géophysique (GMF) du diffusiomètre du satellite ERS-1. Cette méthode est fondée sur les réseaux Multi-Couches. La fonction que nous avons obtenue, NN-GMF, a été étalonnée grâce à la collocation des sigma-zéro mesurés par ERS-1 avec les vecteurs de vent du modèle météorologique Européen ECMWF. Actuellement nous travaillons sur les données mesurées par le diffusiomètre du satellite NSCAT, lancé en juillet 96 par la NASA. A la différence de ERS1 le diffusiomètre de NSCAT dispose de quatre antennes avec des directions différentes. D'autre part, le diffusiomètre envoie les ondes suivant deux polarisations : verticale et horizontale. Ainsi, sous l'hypothèse que les observations sont issues du modèle $\sigma_0 = \sigma_0^*(\vec{v}) + \varepsilon(\vec{v})$ où $\varepsilon(\vec{v})$ représente un bruit blanc gaussien, la modélisation par la méthode du maximum de vraisemblance permet d'estimer différentes sortes de paramètres : moyenne, écart type et covariance (dans le cas de dimensions multiples). Actuellement, nous poursuivons la modélisation afin d'améliorer la fonction de transfert et l'estimation des variances. Nous traitons actuellement, le couple des mesures correspondantes aux deux polarisations horizontales et verticales, pour lesquelles nous estimons la matrice de variance covariance du bruit.

La modélisation par le maximum de vraisemblance permet, en plus de l'estimation des bruits des mesures en sortie, de prendre en compte le bruit des mesures en entrée. Afin de réaliser cette tâche, nous avons défini une nouvelle fonction d'erreur et un algorithme d'apprentissage adapté. Cet algorithme permet, lors de l'apprentissage, d'estimer la fonction de transfert, la matrice de variance covariance du bruit en sortie et de filtrer le bruit en entrée et d'éliminer les points aberrants. Ce travail se fait dans le cadre du contrat européen NEUROSAT.

2 . Problème inverse

Le problème inverse se formule de la manière suivante : étant donnée que la mesure observée est y, quelles sont les valeurs des paramètres x, qui ont inféré l'observation y. Souvent le problème inverse peut être classé comme un problème mal posé. La solution recherchée n'est en général pas unique et le fait que les données soient entachées d'erreurs de mesures accentue l'apparition d'ambiguïtés. Nous avons été confronté à ce type de problème au cours du traitement des données satellitaires, le but étant de déterminer le vent à la surface de l'océan à partir des mesure des trois diffusiomètres embarqués. La détermination de la direction du vent est un problème inverse qui est intrinsèquement ambiguë. Il fallait proposer à une mesure $y = (\sigma_1, \sigma_2, \sigma_3)$ plusieurs directions de vent X avec leur probabilité conditionnelle p(X/y). Il s'agit dans ce cas d'un exemple type d'un problème inverse portant sur des mesures spatiales prises par satellites et qui sont supposées localement indépendantes. La formulation bayesienne de ce problème permet de proposer plusieurs méthodes qui, d'une part, font intervenir un modèle local, et d'autre part, des contraintes globales sur la solution recherchée sous forme d'une probabilité à priori ou d'une solution moyenne acceptable.

- Méthode par modèle local direct utilise la modélisation, par un réseau multicouche, de la fonction de transfert du capteur f() et de la matrice de variance-covariance du bruit, qui seront alors fixées (en gelant les poids du réseau). Cette approche, s'apparente aux méthodes d'inversion par l'adjoint. Le calcul de l'adjoint de la fonction f() par rapport à sa variable d'entrée x peut se faire par simple rétropropagation.

- Méthode par modèle inverse local utilise la modélisation de la fonction densité du paramètre X conditionné par une observation y donnée. Cette fonction densité peut être modélisée par un réseau de neurone qui calcule une mixture de lois normales. Ce type de modèle a été étudié par l'équipe de l'université d'Aston (Royaume Unie) dans le cadre du contrat européen NEUROSAT auquel nous sommes associées.
- Méthode par contexte local, suppose que les paramètres physiques sont spatiallement ou temporellement corrélés, chaque paramètre physique dépend d'un contexte local spatial ou temporel. Nous avons déjà utilisé cette approche en définissant une peudo-vraisemblance pour la reconstitution du champs de vent à partir des mesures du diffusiomètre. D'autre part, nous développons actuellement une méthode qui, sous les hypothèses précédentes, permet de définir une fonction de vraisemblance intégrant le caractère local et global des paramètres.

3 . Cartes topologiques de KOHONEN

L'algorithme des cartes auto-organisatrices de Kohonen est l'un des modèles neuronaux les plus anciens, il représente le prototype de l'apprentissage non supervisé dans le domaine. Un grand nombre de résultats théoriques sont maintenant disponibles, qui explicitent les liens entre ce modèle et les méthodes statistiques de reconnaissance des formes : classification automatique, analyse en composantes principales non linéaires, ... La version dite "batch" des cartes topologiques peut être expliquée à partir du formalisme des nuées dynamiques, formalisme qui permet alors de proposer différentes variantes de l'algorithme initial.

3.1 Interprétation probabiliste des cartes topologiques de KOHONEN

En suivant le formalisme des nuées dynamiques, nous avons proposé un nouvel algorithme (PRSOM) de type carte auto-organisatrices qui permet d'associer à chaque cellule r de la carte une fonction densité gaussienne $f_r(z; w_r, \sigma_r)$ de vecteur moyen w_r et de matrice de variance covariance $\Sigma_r = \sigma_r^2 I$. Nous avons représenté la fonction densité p(z), du générateur des exemples, sous la forme d'une mixture des fonctions densité $p_c(z)$, où $p_c(z)$ est une mixture locale des densités gaussiennes $f_r(z; w_r, \sigma_r)$ définie au voisinage de la cellule c.

L'apprentissage consiste à déterminer les paramètres σ_r et w_r en maximisant la fonction de vraisemblance. Nous avons montré que l'algorithme classique de Kohonen est un cas particulier de PRSOM, ce qui permet donc de donner une interprétation de l'algorithme de Kohonen en terme d'estimation de la fonction densité de l'environnement p(z) par une mixture de fonctions densités.

D'autre part, l'algorithme PRSOM peut être utilisé dans le cadre de l'apprentissage des architectures basées sur les fonctions radiales de bases (RBF). En effet, il constitue une solution intermédiaire entre les deux algorithmes dits "hard" et "soft" et qui sont utilisés classiquement. Cet algorithme dépend d'un hyper paramètre T (de contrôle) qui constitue un paramètre de régularisation, le choix du "meilleur" paramètre T adapté au problème traité, classification ou régression, peut être fait par cross-validation

Nous utilisons actuellement cet algorithme pour faire de la classification avec peu de données expertes (étiquetées).

3.2 Classification avec peu de données expertes

Dans certains problèmes de classification, on dispose d'une quantité énorme de données provenant de mesures à "distance". Vu la taille des données, un expert peut introduire son expertise en étiquetant un nombre limité d'exemples. Il s'agit alors de faire de la classification avec peu de données expertes. Nous développons une méthodologie pour la classification dans ce cadre qui applique la classification hiérarchique aux vecteurs issus de l'algorithme PRSOM. Nous étudions actuellement une famille de mesures de similarité paramètrées et qui contient deux termes : un premier terme qui tient compte de l'inertie des groupes dans l'espace des données et un second terme qui tient compte de la conservation de la topologie dans l'espace discret de la carte. Le choix du paramètre permet de définir la nature du compromis entre ces deux termes. Nous avons appliqué et nous appliquons cette méthode dans des applications réelles.

Classification des nuages. Il s'agit de réaliser une classification des nuages à partir d'images METEOSAT. L'algorithme PRSOM est alors utilisé dans une chaîne de traitement. Les premiers résultats obtenus montrent que PRSOM permet d'obtenir une classification pertinente des nuages. Ce travail a été réalisé dans le cadre du contrat européen NEUROSAT.

Couleur de l'océan. Les capteurs satellitaires mesurant la couleur de l'océan dans une gamme étendue de longueur d'onde permettent d'obtenir des informations intéressantes sur la productivité marine et le contenu en pigments du phytoplancton. Les eaux marines sont divisées quant à leurs propriétés optiques en deux types, les eaux du cas1 où seul le phytoplancton est responsable du signal et les autres. Nous traitons le problème de la classification des spectres de réflectance au sommet de l'atmosphère par une carte topologique de type PRSOM. Les résultats obtenus montrent qu'il est possible d'identifier différents types d'aérosol, ainsi que les différentes classes d'eau (cas-1, cas-2). Cette application est développée actuellement dans le cadre du groupe NEURATEL de l'IPSL dans le cadre du program européen NAOC.

Diagraphie : Elle consiste à faire descendre un appareil de mesure au bout d'un câble dans un trou de forage et à réaliser des enregistrements continus en fonction de la profondeur. Un appareil peut contenir plusieurs sondes mesurant des caractéristiques physiques différentes. Nous traitons deux types de problèmes : la reconstitution de la colonne lithologique et la détermination d'autres paramètres comme la perméabilité et la porosité. Nous avons commencé cette application dans le cadre d'un mémoire d'ingénieur CNAM qui a été financé en partie par la société ELF.

3.3 Classification automatique de données binaires.

Nous étudions actuellement l'aspect conservation de la topologie des cartes topologiques auto-organisatrices ainsi que leur propriété de réduction de la dimension. Notre but est de les adapter aux données qualitatives. Ainsi, en suivant le formalisme des nuées dynamiques, nous avons proposé une méthode de classification automatique de type carte topologique auto-organisatrice adaptée aux données binaires. Nous avons appliqué cette méthode, que nous avons intitulée "bin-batch", à des problèmes classiques afin de la comparer aux algorithmes existants, et nous l'avons appliquée aussi pour la classification des molécules qui sont représentées par des grands vecteurs binaires. Cette dernière application a été traitée dans le cadre d'un stage de DEA qui s'est déroulé à la société L'ORÉAL. Nous développons actuellement une extension à cette méthode dans laquelle on fait intervenir une modélisation sous la forme d'un mélange de loi de Bernouilli. Cette modélisation permet, comme dans le cas de PRSOM, de calculer des probabilités a posteriori d'appartenance aux classes. Cette recherche se fait actuellement dans le cadre de la thèse de M. LEBBAH.

Perspectives.

Nous approfondissons actuellement les points suivants :

1 - Nous continuons notre recherche sur le traitement du problème inverse par des réseaux de neurones. Nous développerons, notamment, les deux méthodes dites d'inversion par modèle direct local et par prise en compte d'un contexte local.

2 -Nous abordons actuellement le problème de l'assimilation des données dans les modèles numériques. Notre but est de définir une méthodologie par des réseaux de neurones permettant d'adapter les paramètres de ces modèles afin qu'ils collent au mieux à la réalité observée. Nous avons entamé cette voie de recherche dans le cadre de deux mémoires de DEA qui ont montré la faisabilité de la méthodologie proposée dans le cadre de l'assimilation des données dans des modèles biologiques en océanographie. Nous continuons cette recherche dans le cadre d'un poste doctorant et qui aura pour objectif de définir les bases pour un développement logiciel de cette méthode.

3 - Nous continuerons nos recherches sur les cartes topologiques auto-organisatrices en nous focalisant sur les points suivants :

- Améliorer la méthode de classification hiérarchique, utilisée pour l'étiquetage des neurones de PRSOM et ceci en définissant de nouvelles mesures de similarités.

- Améliorer la carte de classification des données binaires, en approfondissant la modélisation par un mélange de lois de Bernouilli .

Collaborations scientifiques

Notre équipe a des actions communes avec d'autres groupes de recherches :

- Elle fait partie du groupe NEURATEL (NEURones Appliqués à la TELédétection) et qui regroupe cinq laboratoires de l'IPSL (Institut Pierre Simon Laplace), l'objet de ce groupe est d'appliquer les méthodologies neuronales existantes à des problèmes complexes de télédétection spatiale et de dégager de nouveaux axes de recherche.
- Elle a des collaborations avec M Younes Bennani du LIPN (Laboratoire Informatique de Paris Nord de l'université de Paris 13) ainsi qu'avec Mme Bernadette Dorizi de l'INT d'Evry.
- Elle fait partie, dans le cadre de NEURATEL, du programme européen NAOC (Neural Algorithms for Ocean Color).

Co-Direction de Thèses et de Mémoires.

H DELASSUS. Système de reconnaissance automatique de signatures électromagnétiques en radioastronomie basse fréquence, CNAM, soutenu en mars 99.

M LEBBAH. "Adaptation du contrôle de distance au comportement conducteur", 2eme année de thèse (Université de Versailles, bourse CIFRE).

J. CHEUOUA, Réseaux de neurones et télédétection : restitution des mesures de vecteurs vents à la surface de la mer. Mémoires d'ingénieurs IIE (Institut Informatique d'Entreprise). Ce stage s'est déroulé au LODYC. Mémoire soutenu en juin 99.

M. LEBBAH. Application des réseaux de neurones aux données binaires : Les Cartes Topologiques. Stage de DEA (DEA d'Intelligence Artificielle de l'Université de Paris 13). Ce stage s'est déroulé au département de modélisation et statistique de l'Oréal. Mémoire soutenu en septembre 99.

S. OUIS. Assimilation de données d'un traceur océanique : Une méthodologie Neuronale. Stage de DEA (DEA d'Intelligence Artificielle de l'Université de Paris 13). Ce stage s'est déroulé au LODYC. Mémoire soutenu en septembre 99.

D. FRAYSSINET. Utilisation des réseaux neurones en traitement des données de diagraphies : Prédiction et reconstitution de faciès lithologiques. Mémoire d'ingénieur CNAM soutenu en février 2000.

A. GHOUL. Approche Modulaire pour l'assimilation des données, application à un modèle océanique tridimensionnel de bio géochimie. Stage de DEA (DEA d'Intelligence Artificielle de l'Université de Paris 13). Ce stage s'est déroulé au LODYC. Mémoire soutenu e septembre 2000.

Publications

Articles publiés dans des revues internationales

C. MEJIA, F. BADRAN, A. BENTAMY, M. CREPON, S. THIRIA, N. TRAN (1999) : Determination of the geophysiscal model function of NSCAT and its corresponding variance by the use of neural networks. *J. Geophys. Res*. 1999 special issue , vol 104,NO.C5,pages 11,539-11,556

F BADRAN, M CREPON, C MEJIA, S THIRIA AND N TRAN (1999): Empirical transfert function determination by the use of Multilayer Perceptron. *Neurocomputing Issue* (30) 1-4 pp 31-35.

C AMBROISE, G SEZE, F BADRAN, S THIRIA (1999): Hierarchical Clustering of Self-Organizing Maps for Clouds Classification. *Neurocomputing Issue* (30) 1-4 pp 47-52.

P. RICHAUME, F BADRAN, M CREPON, C MEJIA, H ROQUET, S THIRIA (1999): Neural Network Wind Retrieval from ERS-1 Scatterometer Data. *Neurocomputing Issue* (30) 1-4 pp 37-46

P. RICHAUME, C. MEJIA, S. THIRIA, N. TRAN, M. CREPON, H. ROQUET, F. BADRAN (2000) Neural Network Wind retrieval from ERS1 Scatterometer Data(2000) *J. Geophys. Res*. Vol 105,No C4, pages 8737-8751, April 2000)

Articles publiés dans un ouvrage collectif

M. YACOUB, D FRAYSSINET, F BADRAN, S THIRIA: 2000 "Classification based on Expert knowledge propagation using Probabilistic Self-Organizing Map: application to geophysics" in *Data Analysis : scientific modeling and practical application* Springer-Verlag (Studies in classification, data Analysis, and knowledge organization)

Rapport de recherche

F.BADRAN et S. THIRIA : Les Perceptrons Multicouches: de la régression non-linéaire aux problèmes inverses. Rapport interne CEDRIC et LODYC.

F. BADRAN, Y. STEPHAN, N METOUI, S. THIRIA : A general formulation of non-linear least square regression using multi-layered perceptrons". Rapport interne, CEDRIC et LODYC.

Communications dans des congrès avec comité de sélection et actes

FRAYSSINET, D., S. THIRIA, F. BADRAN, and L. BRIQUEU: Use of neural networks in log's data processing: prediction and rebuilding of lithologic facies. Petrophysics meets Geophysics, Paris, 2000.

N. TRAN, S. THIRIA, M. CREPON, F. BADRAN and M.FREILICH: 'Validation of the QSCAT NRCS on the advanced neural network NSCAT GMF and estimation of neural network QSCAT GMF'. IGARSS'2000, Honolulu Hawai, July 24-28, 2000.

M LEBBAH, F BADRAN, S THIRIA (2000): Topological Map for Binary data. *ESANN'2000, the European Symposium on Artificial Neural Networks*. April 26-27-28, 2000. Bruges, Belgium.

METOUI, F. BADRAN (1999). Regression non lineaire par perceptrons multicouches. *XXXIemes Journées de Statistique* ,17-21 mai 1999, Grenoble.

M LEBBAH, F BADRAN, S THIRIA (2000). Carte Toplogique et données binaires. *XXXII ème journées de Statistique*. FES 15-19 Mai 2000.

M. YACOUB, F. BADRAN, S. THIRIA, (2000). Cartes topologiques et classification hiérarchique : application a la classification de l'océan. *XXXII ième journées de Statistique*. FES 15-19 Mai 2000.

Communications soumises à des congrès (2001)

M. LEBBAH, C. CHABANON, S.THIRIA, and F. BADRAN (2001). Probabilistic Topological Map and Binary data. Soumis à ICANN2001.

M. YACOUB, N. NIANG, F. BADRAN, and S. THIRIA (2001). A New Hierarchical Clustering Method using Topological Map. Soumis à ASMDA2001

M. Yacoub, F. Badran, and S. Thiria.. A Topological Hierarchical Clustering: Application to Ocean Color Classiffication. ICANN90.