

Rôle du contexte dans le modèle OWE*

- un réseau de neurones artificiels -

- utilisant des connexions axo-synaptiques -

Laurent Bougrain¹, Nicolas Pican², Frédéric Alexandre¹

¹ LORIA - INRIA Lorraine, Bâtiment LORIA, Campus scientifique B.P. 239,
54506 Vandœuvre-lès-Nancy Cedex, France - E-mail: bougrain@loria.fr, falex@loria.fr

² Computing & Electrical Engineering Dpt, Heriot-Watt University
Riccarton, EH14-4AS Edinburgh, Scotland. E-mail: ceenvp@cee.hw.ac.uk

Résumé : Nous avons établi un modèle neuronal contextuel à partir de données neuro-biologiques soulignant le rôle modulateur de certains neurones sur l'apprentissage d'autres neurones. Aujourd'hui, ce modèle est assez éprouvé pour être utilisé simplement et efficacement sur des applications industrielles de vraie grandeur. Nous reportons ici une de ces applications. Nous exploitons également ces résultats afin de tenter de répondre à la question du choix des paramètres pouvant être considérés comme contextuels dans de tels problèmes multidimensionnels.

1 Le principe de l'OWE

Dans les réseaux de neurones artificiels, les connexions entre neurones sont généralement de type axo-dentritique. La connexion se fait par l'intermédiaire des terminaisons synaptiques de l'axone d'un neurone afférent j sur les dendrites d'un neurone i . L'activation du neurone j est alors pondérée par le poids de la connexion w_{ij} et transmise au neurone i (voir Figure 1).

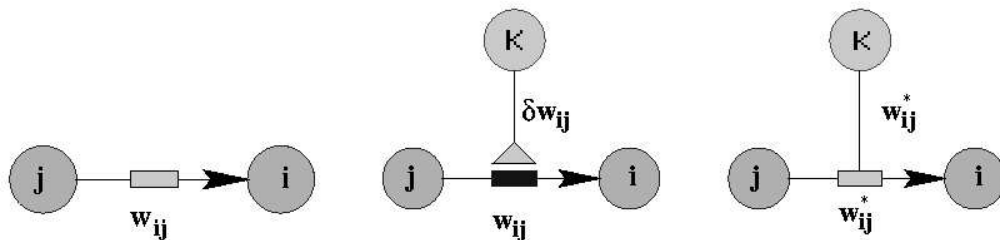


FIG. 1 –. connexions axo-dentritique classique, et axo-synaptique de ODWE et de OWE

Des recherches en neurobiologie ont montré qu'il existe d'autres types de connexions entre neurones tels que: axo-somatique (transmission directe de l'activité pré-synaptique au soma), axo-axonique (transfert d'activité post-synaptique à un autre axone) et axo-synaptique (connexion d'une synapse sur une autre synapse modulant ainsi son efficacité) [3] [2] .

Nous avons montré [6] qu'une connexion axo-synaptique pouvait se modéliser par le principe de l'ODWE (Orthogonal Delta Weight Estimator). De façon synthétique, la structure d'une connexion ODWE peut se représenter par trois neurones (i, j, k) avec une connexion axo-dentritique S_{ij} entre

* Ce travail a bénéficié du soutien du CNET au titre du marché n97 1B008

j et i et la terminaison synaptique du neurone k connectée sur la synapse S_{ij} (voir Figure 1). Le principe consiste à augmenter la probabilité de libération de neuromédiateurs de la synapse S_{ij} dans son espace synapto-dendritique par l'action du neurone k appelé neurone modulateur. Le calcul de l'activation du neurone i peut alors se traduire par l'équation suivante:

$$Y_i = F_i((w_{ij} + \delta w_{ij})Y_j) \quad (1)$$

qui se généralise par:

$$Y_i = F_i\left(\sum_j (w_{ij} + \delta w_{ij})Y_j\right) \quad (2)$$

dans laquelle δw_{ij} représente la modulation (sortie du neurone modulateur k), w_{ij} le poids de la connexion axo-synaptique entre j et i supposé figé (ie le poids statique de la synapse S_{ij}), Y_i et Y_j l'activation des neurones i et j et F_i la fonction de transfert du neurone i .

Le principe de l'OWE (Orthogonal Weight Estimator) utilise une variation de la modélisation d'une connexion axo-synaptique définie pour l'ODWE. Le poids effectif de la synapse S_{ij} , qui est dans le cas de l'ODWE: $w_{ij} + \delta w_{ij}$ avec w_{ij} figé, devient pour l'OWE w_{ij}^* , où w_{ij}^* est égal à la sortie du neurone "modulateur" k (voir Figure 1).

Ce modèle généralise en fait le modèle d'unités $\Sigma\Pi$, où certaines afférences ont une influence multiplicative [1] et par là même modulatrice sur l'activité des unités. Ainsi, l'intérêt des principes OWE et ODWE est d'obtenir une architecture capable d'apprendre et de modéliser des comportements contextuels (voir Figure 2). En effet, si le neurone modulateur est le neurone de sortie d'un MLP (appelé RN OWE ou ODWE) ayant en entrée des paramètres représentant le contexte d'une perception et si les neurones i et j appartiennent à un réseau de neurones (appelé RN principal) ayant en entrée une perception, les sorties du RN principal seront fonction des perceptions, mais adaptées à leur contexte.

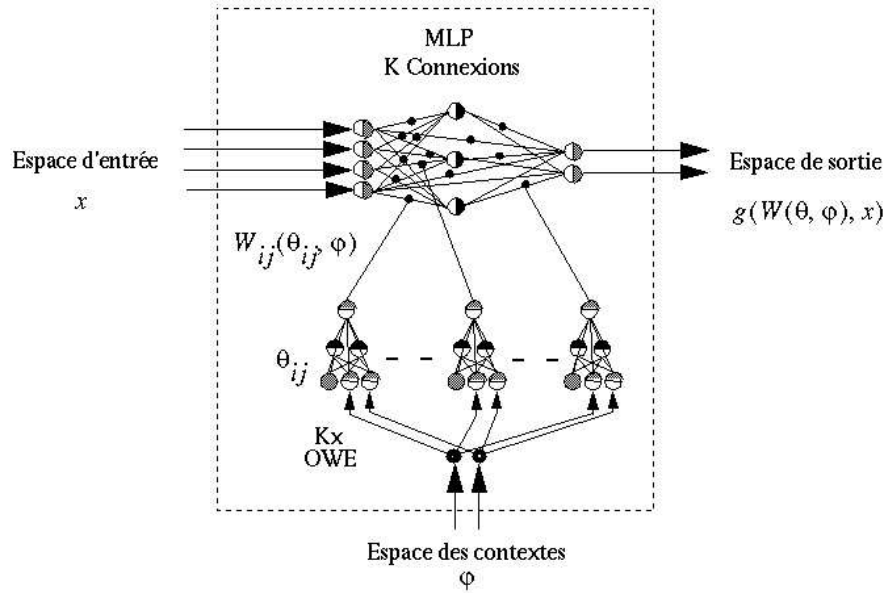


FIG. 2 –. architecture de OWE

Le lecteur trouvera toutes les informations sur les architectures OWE et ODWE dans [6] [7] ainsi que des applications en contrôle adaptatif [5] et en reconnaissance de la parole continue [4]. Nous rapportons ci-dessous une autre application pour ce modèle qui servira à observer l'influence du contexte sur les résultats.

2 Le problème contextuel de prédiction du champ radioélectrique

L'application choisie se situe dans le domaine de la prédiction de champs radioélectriques pour l'optimisation des systèmes de communication. Ce problème est fortement contextuel puisqu'il dépend de la nature de l'environnement. Sa complexité nécessite la prise en compte de paramètres spécifiques à la zone d'étude. Afin de prendre en compte l'environnement dans l'estimation du champ radioélectrique, nous utilisons une base de données du CNET comprenant 32 paramètres. Cette base contient une description du sursol entre l'émetteur et le récepteur ainsi que des données sur le positionnement relatif de la source par rapport à la cible. Les corpus d'apprentissage et de test font partie d'une validation croisée et contiennent le même nombre d'exemples soit 10016.

2.1 Le perceptron multicouches comme référence

Dans un premier temps, ces paramètres seront présentés en entrée à un perceptron multicouches (MLP) dont les vertus d'estimateur universel sont connues. Les résultats nous donneront une valeur de référence que l'on comparera à celle de OWE. L'architecture utilisée peut être décrite sous forme (32x10x1) où les trois valeurs correspondent respectivement au nombre de neurones de la couche d'entrée, cachée et de sortie.

De plus, si l'on effectue un partitionnement des données à l'aide d'un algorithme non supervisé comme Growing Neural Gas, après apprentissage d'un MLP pour chaque partition (soient 9 MLPs de type (32x10x1)), nous avons une confirmation de la composante contextuelle du problème. En effet, l'écart type des valeurs estimées par rapport aux valeurs désirées diminue fortement (voir les résultats des multi-MLPs dans le tableau 3. Ces derniers représentent la moyenne des valeurs absolues des erreurs, la somme des erreurs et l'écart type des erreurs).

2.2 L'apport de OWE

Jusqu'à présent, nous avons vérifié l'intérêt d'une approche contextuelle mais sans connaissance précise des paramètres pertinents pour caractériser ce contexte. Les différentes méthodes utilisées pour la classification sont non supervisées et ne permettent pas d'extraire un lot de paramètres. Tous les paramètres sont d'une certaine manière contextuels sinon ils n'auraient pas été retenus après l'application de techniques d'élagage. OWE sera d'autant plus performant que le contexte sera adapté au problème.

Dans un premier temps, le choix des paramètres contextuels se fera d'après leur nature en accords avec les experts du domaine. On peut faire varier le nombre de paramètres utilisés, d'un nombre restreint (8) à la quasi totalité (28), pour étudier le nombre de paramètres pertinents qui décriront le contexte dans le modèle OWE. Les résultats (tableau 3) montrent que le contexte est mieux caractérisé par un petit nombre de paramètres bien choisis. De plus, non seulement la prédiction est meilleure que celle d'un large contexte mais le modèle OWE couplé à un MLP améliore la performance d'un simple MLP.

D'autre part, nous voulons vérifier l'hypothèse que les paramètres les plus corrélés à la valeur à prédire sont de bons paramètres contextuels. Puisque nous avons vu qu'un petit nombre de paramètres pouvaient constituer un contexte pertinent pour OWE, seuls les 7 paramètres les plus corrélés à la valeur à prédire ($|\text{corrélation}| > 0.56$) et les 9 paramètres les moins corrélés ($|\text{corrélation}| < 0.074$) seront utilisés pour définir les deux contextes de test. Les résultats confirment que les paramètres fortement corrélés à la valeur à prédire constituent un meilleur choix de contexte pour le modèle OWE que les valeurs subjectives choisies par les experts. De plus, les paramètres peu corrélés sont un très mauvais choix. Ils ne permettent pas d'obtenir une bonne spécialisation du modèle d'inspiration biologique OWE.

Modèle	Architecture	Phase	Moyenne	Somme	Ecart type
<i>MLP</i>	(32x10x1)	<i>Apprentissage</i>	4.16	0.0002	5.28
		<i>Test</i>	4.18	0.22	5.39
<i>Multi-MLPs</i>	9 MLPs (32x10x1)	<i>Apprentissage</i>	3.61	0.0002	4.64
		<i>Test</i>	3.72	-0.29	4.76
<i>OWE+MLP (petit contexte)</i>	(8x8x1)+(24x8x1)	<i>Apprentissage</i>	3.82	-0.0003	4.84
		<i>Test</i>	4.07	-0.53	5.23
<i>OWE+MLP (large contexte)</i>	(28x6x1)+(4x4x1)	<i>Apprentissage</i>	28.72	0.0001	36.79
		<i>Test</i>	32.88	-4.12	42.27
<i>OWE+MLP (contexte fortement corrélé)</i>	(7x4x1)+(25x10x1)	<i>Apprentissage</i>	3.56	0.0003	4.55
		<i>Test</i>	4.01	0.29	5.18
<i>OWE+MLP (contexte faiblement corrélé)</i>	(9x4x1)+(23x10x1)	<i>Apprentissage</i>	29.57	-0.0007	37.75
		<i>Test</i>	34.15	-5.61	43.93

FIG. 3 –. Tableau des erreurs de l'estimation

3 Conclusion

Issu à l'origine d'un modèle biologique, l'algorithme d'apprentissage OWE est maintenant utilisé dans de vraies applications industrielles. On montre clairement que, par l'utilisation de ce modèle, les performances de réseaux classiques (MLP) sont améliorées, et qu'elles s'approchent de celles obtenues avec une architecture Multi-MLPs. Par contre, deux avantages sont en faveur de l'utilisation des OWEs: ils ne nécessitent qu'une seule phase d'apprentissage et incluent une faculté d'interpolation entre les contextes. Suite à quoi, il nous reste à mieux définir les critères permettant de définir objectivement quels paramètres d'un problème doivent être considérés comme étant contextuels. Les résultats reportés ici vont dans le sens d'une meilleure connaissance de ces critères et devraient être approfondis pour mieux les appréhender.

Références

1. F. Alexandre and F. Guyot. Neurobiological inspiration for the architecture and functioning of cooperating neural networks. In *Proceedings of Int. Workshop on ANNs*, Malaga, Spain, June 1995.
2. C. H. Bailey M. Chen. Morphological basis of long term habituation and sensitization in aplysia. *Science*, 220:91–93, 1983.
3. J. Delacour. Apprentissage et mémoire : une approche neurobiologique. *Masson (Ed.)*, 1987.
4. N. Pican D. Fohr J-F. Mari. Continuous speech recognition using a context sensitive ann and hmms. *EUROSPEECH'97 Proceedings*, 1997.
5. N. Pican. Approches statique et dynamique de la modulation des efficacités synaptiques dans les réseaux de neurones. *PhD. Thesis*, 1995.
6. N. Pican. An orthogonal delta weight estimator for mlp architectures. *ICNN'96. Washington, DC, USA*, 1996.
7. N. Pican. Contextual kohonen som with orthogonal weight estimator principle. *ICANN'97 Proceedings*, 1997.