Les mathématiques de la mémoire

Aude Forcione-Lambert

Université de Montréal, Département de mathématiques et de statistique

Introduction

En 1940, le monde scientifique vit une révolution technologique : les «machines à calculer» viennent de troquer le rouage pour le transistor. Les neurologues ne tardent pas à faire le parallèle entre ce nouvel outil - l'ordinateur - et le cerveau. Naissent alors deux domaines frères : la neurologie computationnelle, cherchant à comprendre le système nerveux par la modélisation, et l'intelligence artificielle, cherchant plutôt à égaler les performances de l'humain.

Durant les 30 dernières années, les modèles de part et d'autre ont explosé en complexité, requérant des outils d'analyse de plus en plus poussés afin de comprendre leur fonctionnement. Les mathématiciens spécialisés en systèmes dynamiques se taillent donc maintenant une place de choix dans le domaine.

Résumé

Notre recherche se penche sur un réseau de neurones devant implémenter des tâches simples de mémoire, et plus précisément sur l'analyse de la solution trouvée par l'algorithme d'optimisation. Nous prenons comme point de départ une étude de Sussillo et Barak[1]. Le fonctionnement de notre réseau de neurones est significativement différent de celui utilisé dans l'article, mais il doit accomplir la même tâche et est soumis aux mêmes analyses. Par la suite nous expérimentons avec une tâche modifiée et observons comment le réseau ajuste sa stratégie en fonction des nouveaux objectifs. Les résultats soulèvent des questions intéressantes à la fois en neurologie et en IA.

Tâche

La tâche originale consistait à garder en mémoire la dernière entrée pour chacun des canaux. La nouvelle tâche ajoutait un délais entre l'entrée et la sortie.

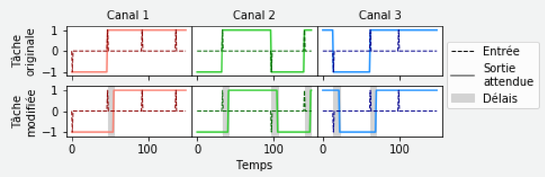

Figure – Exemple de couple entrée-sortie pour la tâche otiginale et la tâche modifiée.

Modèle

On utilise un réseau de neurones à pas de temps discret avec une unique couche récurrente. Le vecteur d'entrée X et de sortie Y sont chacun de longueur 3 alors que le vecteur S représentant l'état des neurones est de longueur 100. La matrice W_{in} représente les connections entre les entrées et les neurones, W_{out} entre les neurones et la sortie et W est une matrice 100×100 qui représente les connections récurrentes entre les neurones. On met à jour le système d'après l'équation suivante :

$$S(t) = \tanh(X(t)W_{in} + S(t-1)W)$$
$$Y(t) = \tanh(S(t)W_{out})$$

Le modèle utilisé par Sussillo et Barak était semblable mais le temps était continu, avec une équation différentielle qui en définissait la dynamique.

Outils mathématiques

Analyse en composantes principales (ACP) : Pour mieux visualiser la dynamique, on aimerait trouver le sous-espace de l'espace de phase qui contient le plus d'informations. On diagonalise la matrice de covariances en ordonnant les valeurs propres de façon ascendante :

$$C_{diag} = PCP^T$$

La matrice P donne la base dont les premières dimensions contiennent le plus d'informations (c'est à dire sur lesquelles les données ont la plus grande variance). Ce sont les composantes principales (CPs).

Analyse des points lents : Les points fixes donnent des informations importantes sur la dynamique d'un système. Comme trouver les points fixes numériquement est difficile, on s'inspire de la physique pour créer un champ scalaire dont les minimums sont les «points lents» :

$$q = \frac{1}{2}(S(t+1) - S(t))^2$$

Résultats

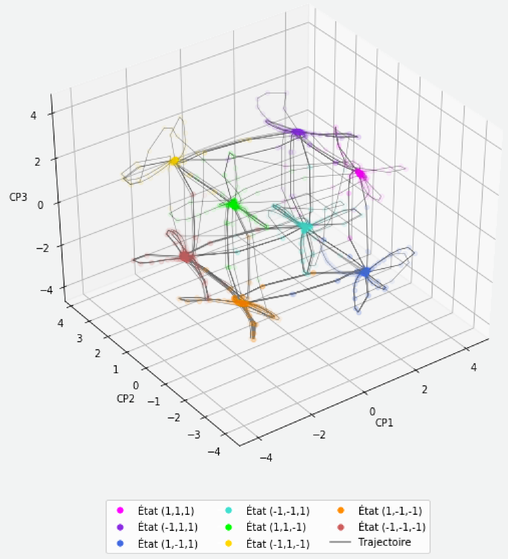

Figure – Dynamique durant 3000 itérations pour la tâche originale selon les trois premiers CPs.

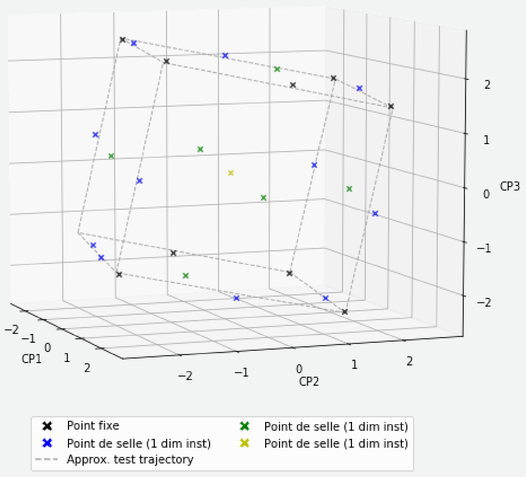

Figure – Points fixes pour la tâche originale selon les trois premiers CPs. Le cube en traits pointillés représente la dynamique approximative durant 3000 itérations.

Le modèle implémente la tâche originale grâce à des attracteurs sur les sommets d'un «cube». Des points de selle canalisent les changements d'état.

Résultats

La tâche modifiée est implémentée par un processus semblable à une transformée de Fourier dont les éléments sont amortis.

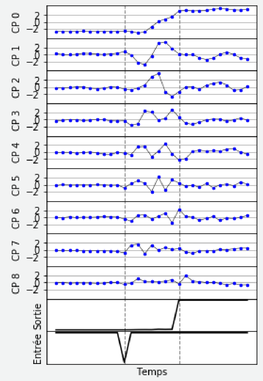

Figure – Dynamique d'une transition selon les 9 premiers CPs pour un délais de 8 itérations.

Conclusion

Nos observations quant à l'implémentation de la tâche originale sont identiques à celles de Sussillo et Barak. Or, notre notre modèle était en temps discret à l'opposé du leur en temps continu. De plus, les algorithmes d'entraînement étaient différents. Cela suggère une certaine universalité de la stratégie d'implémentation optimale.

L'implémentation de la tâche modifié est un bon exemple d'émergence de schémas temporels dans un système neuronal. Plusieurs recherches récentes ont démontré l'utilité de ces shémas dans la réalisation de tâches complexes, autant dans le cerveau humain que dans les réseaux de neurones artificiels.

Références

[1] David Sussillo and Omri Barak.
Opening the black box : low-dimensional dynamics in high-dimensional recurrent neural networks.
Neural computation, 25(3) :626–649, 2013.

Remerciements

Merci à Dr.Guillaume Lajoie pour avoir dirigé ce projet.

Mes coordonnées

- Web : github.com/aude-forcione-lambert
- Courriel : aude.forcione.lambert@gmail.com