

1 Introduction

Le cerveau est amené dans certaines situations à déduire les causes des sensations comme par exemple la sensation du froid qui est liée à la présence d'un courant d'air. Nous montrons dans ce résumé que le cerveau a l'équipement anatomique et physiologique nécessaire pour inverser les modèles mathématiques qui lient les causes des sensations aux signaux sensitifs. Cela permet au cerveau d'émettre des hypothèses sur son environnement à partir des informations que lui transmettent ses sens. L'article s'intéresse dans un premier temps aux modèles dynamiques hiérarchiques et à leur inversion, ensuite il montre comment l'inversion peut être formulée comme une simple ascension de gradient en utilisant des réseaux neuronaux, et enfin il détaille comment les réponses cérébrales peuvent être comprises comme des interférences dans les modèles dynamiques hiérarchiques.

2 Modèles hiérarchiques dynamiques

Les modèles étudiés sont supposés être de la forme :

$$y = g(x, v, \theta) + z$$

$$\dot{x} = f(x, v, \theta) + w$$

avec v la source à l'origine de la sensation, y les données reçues par le cerveau (la sensation), x l'influence de v sur y , θ l'ensemble des paramètres extérieurs au cerveau, z et w du bruit, et f et g sont des fonctions non linéaires.

En dérivant successivement ces équations, on obtient :

$$\tilde{y} = \tilde{g} + \tilde{z}$$

$$D\tilde{x} = \tilde{f} + \tilde{w}$$

où $\tilde{y} = [y, y', y'', \dots]^T$ et D est une matrice par bloc représentant l'opérateur dérivation.

Ce modèle se généralise en un modèle dynamique hiérarchique en liant les différentes équations, et en supposant que v suit une loi gaussienne de moyenne η :

$$\begin{aligned}
 y &= g(x^{(1)}, v^{(1)}) + z^{(1)} \\
 \dot{x}^{(1)} &= f(x^{(1)}, v^{(1)}) + w^{(1)} \\
 &\dots \\
 v^{(i-1)} &= g(x^{(i)}, v^{(i)}) + z^{(i)} \\
 \dot{x}^{(i)} &= f(x^{(i)}, v^{(i)}) + w^{(i)} \\
 &\dots \\
 v^{(m)} &= \eta + z^{(m+1)}
 \end{aligned}$$

Ce système est défini hiérarchiquement : la sortie de chaque niveau est l'entrée du suivant, et il y a de nombreux niveaux cachés entre la source initiale et la perception du cerveau. Il permet de rendre compte de la structure particulière de ces modèles.

L'objectif est maintenant de l'inverser, pour pouvoir retrouver la source initiale $v^{(m)}$ à partir de la sensation y . Pour cela, la méthode variationnelle de Bayes permet d'avoir une approximation de la probabilité conditionnelle $p(\{x; v; \theta\} | y, m)$, connaissant les données y et le modèle m . Cela est possible en maximisant la densité $q(\{x; v; \theta\})$ connaissant une borne inférieure de $p(y|m)$. Avec $q(\{x; v; \theta\}) = q(u(t))$ où $u(t) = [\tilde{v}, \tilde{x}]^T$, l'approximation de Laplace permet d'estimer $q(u(t)) = N(u : \tilde{\mu}, C)$ où N est la loi normale, $\tilde{\mu}$ la moyenne conditionnelle et C la matrice de covariance. Il s'agit alors d'estimer $\tilde{\mu}$, sachant que la matrice de précision du bruit (l'inverse de la matrice de covariance, notée $\tilde{\Gamma}$) est une fonction de $\tilde{\mu}$.

Dans un système dynamique, la moyenne maximise l'intégrale de chemin de l'énergie $U(t) = \ln(p(\tilde{y}, u))$, ce qui revient à résoudre l'équation :

$$\dot{\tilde{\mu}} - D\tilde{\mu} = U(t)_u$$

Ces modèles dynamiques ont donc une structure hiérarchique particulière, et l'inverser peut revenir à résoudre l'équation ci-dessus.

3 Implémentation neuronale des modèles hiérarchiques dynamiques

3.1 L'architecture neuronale biologique

Le cerveau a une organisation hautement hiérarchisée qui correspond très bien au modèle présenté en 2. Cette hiérarchie (distinction entre les couches "hautes" et "basses" du réseau) semble difficile à établir au premier abord à cause de l'omniprésence de boucles de rétro-action, mais des différences biologiques fondamentales existent entre les connexions vers l'avant et celles vers l'arrière et qui permettent de les distinguer :

- leur forme : par exemple, les neurones des couches supra granulaires connectées vers l'avant du réseau sont pyramidales superficielles, tandis que celles connectées vers l'arrière sont pyramidales profondes,
- leur fonctionnement physique et chimique : il diffère en fonction de leur but, par exemple les réactions sont plus lentes dans le cas d'une rétropropagation de signal.

Cela semble impliquer que les signaux propagés vers l'avant constituent des ordres directs appliqués rapidement, tandis que ceux propagés vers l'arrière agissent pour améliorer la qualité de ces ordres sur un plus long terme.

3.2 Dédution des causes d'une sensation

Dans un précédent article, les auteurs ont démontré que les erreurs de prédiction du cerveau sur les causes d'une sensation s'écrivent sous la forme suivante, qui est aussi une expression de la dynamique des neurones :

$$\begin{aligned}\varepsilon^v &= \begin{pmatrix} y \\ v^{(1)} \\ \vdots \\ v^{(m)} \end{pmatrix} - \begin{pmatrix} g^{(1)} \\ g^{(2)} \\ \vdots \\ \eta \end{pmatrix} \\ \varepsilon^x &= \begin{pmatrix} Dx^{(1)} \\ \vdots \\ Dx^{(m)} \end{pmatrix} - \begin{pmatrix} f^{(1)} \\ \vdots \\ f^{(m)} \end{pmatrix} \\ \tilde{\varepsilon} &= \begin{pmatrix} \tilde{\varepsilon}^v \\ \tilde{\varepsilon}^x \end{pmatrix}\end{aligned}$$

On obtient alors un modèle mathématique de l'évolution des neurones lors de la réception d'un signal : $\dot{\tilde{\mu}} = U(t)_u + D\tilde{\mu} = D\tilde{\mu} - \tilde{\varepsilon}_u^T \xi$, avec $\xi = \tilde{\Pi} \tilde{\varepsilon}$ et $\tilde{\Pi} = \begin{pmatrix} \tilde{\Pi}^z \\ \tilde{\Pi}^w \end{pmatrix}$. Ce modèle fait ressortir clairement deux types de neurones différents que l'on peut retrouver biologiquement (voir section précédente) : les neurones de propagation (qui encodent $\tilde{\mu}$) et les neurones de rétropropagation d'erreur (qui encodent ξ , qui représente l'erreur de prédiction pondérée par la précision). Par ailleurs, les équations montrent également que les différents types de neurones s'envoient des messages les uns aux autres de sorte à optimiser leurs réponses.

3.3 La transmission de messages au sein du réseau

En développant les équations précédentes, on trouve que les neurones de rétropropagation reçoivent des messages provenant de leur niveau et du niveau directement inférieur. De même, les neurones de propagation reçoivent des messages de leur propre couche et de celle directement supérieure (voir figure ci-dessous). Par conséquent, toutes les interactions entre deux couches sont réciproques, ce qui signifie qu'entre deux couches 1 et 2 n'existent comme interactions que la propagation de 1 vers 2 (qui représente le signal de la prédiction faite sur l'origine de la sensation) et la rétropropagation de 2 vers 1 (qui représente l'erreur effectuée sur cette prédiction). Toutes ces conséquences du modèle mathématiques présenté ci-dessus peuvent être vérifiées dans le comportement des cellules correspondantes. Le cerveau dispose donc bien de la structure nécessaire à la réalisation de l'opération d'inversion de signal décrite dans la première partie de l'article.

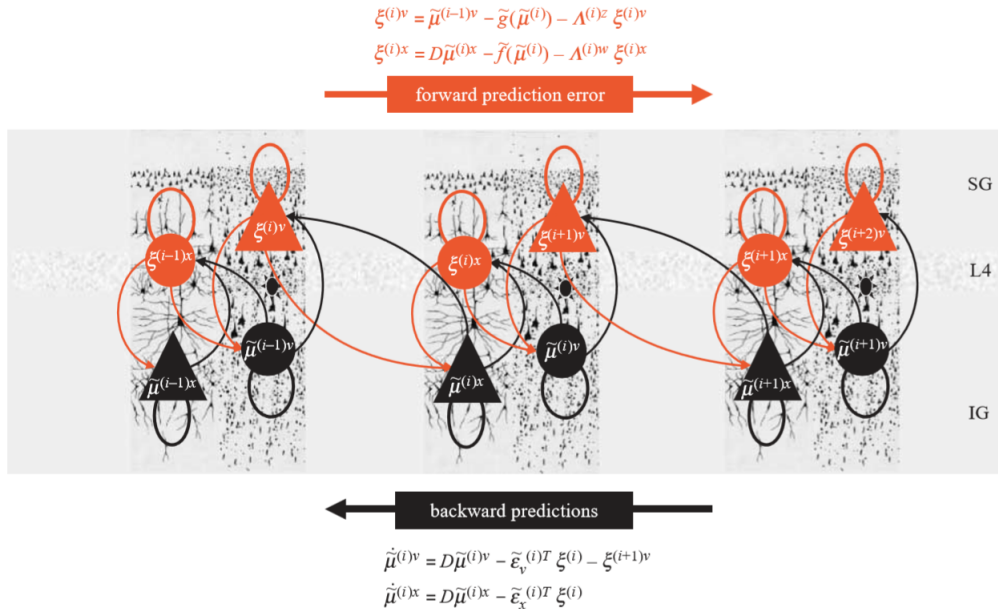


FIGURE 3.1 – Structure du réseau neuronal avec les différents sens des connexions

4 Application aux chants d'oiseaux

Afin d'examiner les propriétés d'un réseau neuronal issu d'un modèle hiérarchique dynamique, les auteurs de l'article utilise des méthodes non-invasives telles que la mesure du potentiel local LFP (Local Field Potential) ou ERP (Evoked Response Potential). Les cellules pyramidales sont adapté à ce type de mesures, et on assigne ces signaux (LEP et ERP) à l'erreur de prédiction.

Les signaux reçus par le cerveau au cours du temps modifient ses cellules neuronales. Ce système dynamique crée des attracteurs au sein du réseau qui évoluent avec les signaux vers une prédiction sur l'environnement. Se définissent ainsi des couches d'attracteurs superposées. Ce modèle d'attracteurs a des implications sur la structure temporelle de la perception. La dynamique des représentations de haut niveau se déroule plus lentement que la dynamique des représentations de niveau inférieur, ce qui corrobore les précédentes constatations sur les propriétés biologiques des neurones en fonction de leur rôle (voir partie 3).

L'article s'intéresse par la suite à l'exemple du chant d'oiseau et fait une série d'expériences pour confirmer la théorie.

Certes beaucoup d'efforts de modélisation utilisant des attracteurs ont été fait en biomécanique pour comprendre la génération de chant d'oiseau mais dans le cas présenté dans cet article, on utilise ces attracteurs pour contrôler les sonogrammes au cours du temps. Ce modèle est basé sur deux attracteurs de Lorenz : un premier contrôlant la fréquence et l'amplitude du signal, et un deuxième permettant d'influencer le premier, afin de créer une structure hiérarchique. L'objectif est de tester le modèle construit plus haut afin de voir s'il est capable d'identifier des chants d'oiseau.

On constate que les paramètres structurels et dynamiques du chant sont nécessaires pour permettre au modèle d'anticiper la suite du chant qu'il entend. Lorsque l'on supprime la fin de ce chant, l'erreur de prédiction augmente considérablement, et est entièrement générée par la prédiction qu'a faite l'oiseau, puisqu'il n'y a plus de signal reçu. De plus, on observe un signal perçu après la coupure du chant, qui correspond à celui qui aurait été perçu s'il n'y avait pas eu cette coupure. Cette expérience rend bien compte des capacités de prédiction du cerveau. Il est intéressant de noter que la suppression des paramètres structurels du chant a un effet moins marqué sur la reconnaissance que la suppression des paramètres dynamiques.

Dans les simulations précédentes, il a bien été montré que les chants sont des suites de notes ordonnées par un attracteur défini par un ensemble de paramètres : à chaque point dans l'espace des états d'un attracteur correspond une catégorie de chant. Même si le chant perçu n'est pas exactement reconstitué à la perfection, la précision est suffisante pour éviter les confusions. Cet exemple montre que le modèle considéré est suffisamment complexe pour qu'il permette de remonter d'un signal à une prédiction.

5 Conclusion

Pour conclure, cet article a bien montré que lors de son interprétation des signaux sensoriels, le cerveau suit des modèles mathématiques proches de l'inversion d'un système dynamique hiérarchique. Il réalise cette inversion à l'aide de signaux de deux types qui se propagent dans son réseau de neurones : une propagation vers l'avant d'un signal de prédiction et une rétropropagation de l'erreur commise sur cette prédiction. Ces réseaux évoluent alors au cours du temps vers des attracteurs, qui peuvent être interprétés comme les catégories dans lesquelles le cerveau classe les signaux qui lui parviennent. Ces hypothèses sont corroborées par des simulations de chants d'oiseau qui montrent que ce modèle permet bien une classification précise de signaux étendus dans le temps.