

Apprentissage par imitation

S. Moga & P. Gaussier

Groupe Neurocybernetique, ETIS URA CNRS 2235,

6 Avenue du Ponceau, 95014 Cergy, France.

e-mail: gaussier ou moga@ensea.fr

Résumé

Nous proposons une architecture neuronale qui permet à un robot mobile (*l'élève*) d'apprendre à imiter une séquence de mouvements exécutés par un autre robot ou un humain (*le professeur*). Le réseau de neurones utilisé pour l'apprentissage de séquences est directement inspiré de la modélisation d'une structure du cerveau, l'hippocampe.

1. Introduction

Les travaux réalisés dans notre équipe sont orientés vers la réalisation d'une architecture neuronale de contrôle pour un robot autonome guidé par la vision. Utilisant une architecture Perception-Action (PerAc) [4] nous avons montré [3] que, dans un milieu ouvert, notre robot peut atteindre une position particulière avec une grande précision en utilisant seulement des amers visuels. Mais la découverte des lieux intéressants pour une tâche particulière est un problème NP-complet (le temps d'apprentissage augmente très vite avec la taille du problème). En conséquence, l'utilisation de l'imitation pour apprendre un comportement, ou une partie d'un comportement, est une approche qui peut permettre à une population de robots d'apprendre à trouver des solutions par eux-mêmes. L'apprentissage par imitation est déjà utilisé dans quelques projets d'intelligence artificielle (pour une présentation générale de tous ces travaux voir [1]). Contrairement à ces approches, notre architecture permet un apprentissage en ligne et non supervisé.

2. Imitation

L'idée de notre approche est simple : l'imitation, dans sa forme la plus simple, peut être déclenchée par une erreur de perception. Prenons, par exemple, le cas d'un bras de robot (A). On suppose que le robot a déjà appris les associations entre la position visuelle de son bras (information visuelle) et les angles de ses articulations (information motrice). Supposons maintenant que la caméra est orientée dans la direction d'un autre bras de robot (B). Parce que le premier bras de robot (A) essaie de corriger la différence entre la représentation visuelle de ce qu'il croit être son bras (information visuelle) et la représentation de la position de son bras (information motrice), il fait les mêmes mouvements que le deuxième bras (B).

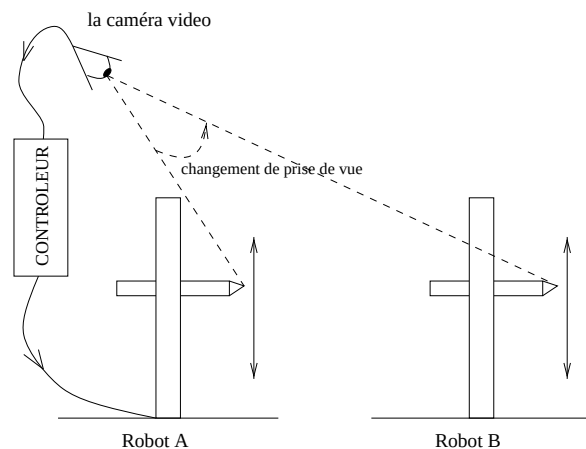


Figure 1: Un observateur extérieur pourra déduire qu'il est en train d'imiter.

Si la séquence de mouvements du bras (A)

mouvements peut être apprise et associée à une motivation particulière.

Dans notre problème le robot élève essaie de réduire la différence entre la représentation de sa perception visuelle du mouvement et celle de son activité motrice (le principe d'homéostasie). L'intégration de tous les systèmes (d'apprentissage, homéostatique, de prédiction, motivationnel) est faite à travers d'une architecture neuronale de type PerAc. Dans cette architecture le système homéostatique est implémenté comme un lien réflexe. Il déclenche l'action motrice qui permet d'annuler le flot visuel créé par les mouvements du *professeur*. Le mécanisme d'apprentissage permet par ailleurs les associations entre les actions motrices et le système motivationnel.

3. Suivi d'un autre robot

A partir d'une séquence d'images prises par la caméra (CCD dans fig. 4) embarquée sur le robot nous fabriquons une image du mouvement perçu (fig. 2a). L'image du mouvement est directement copiée sur une carte de neurones (MI dans fig. 4). Ensuite, grâce à des mécanismes de compétition et de projection (qui correspondent à des neurones avec un large champ récepteur) la direction et la vitesse du mouvement sont extraites (fig. 2b). Par la suite un changement d'orientation et/ou de vitesse est proposé à la carte de neurones moteurs (MO dans fig. 4).

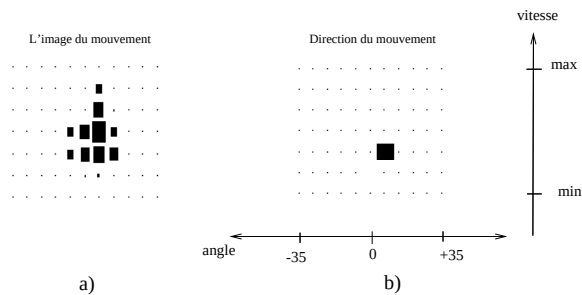


Figure 2: a) Le flot visuel créé par les mouvements du *professeur*. b) La direction et la vitesse proposées pour le mouvement de *l'élève*.

Dans un premier test, nous avons connecté

visuel à la carte de mouvements. Le suivi du professeur était correct mais les ordres moteurs étaient très bruyés. Ce bruit était le résultat du fait qu'il existe toujours une incertitude sur la position du centre de la zone en mouvement perçue par *l'élève*. L'utilisation de mécanismes classiques de filtrage élimine ce problème mais rend délicate la segmentation des mouvements du *professeur* (transitions trop lissées).

Nous avons développé un mécanisme qui découple la direction de la tête du robot de la direction de son corps en lui donnant la possibilité de faire tourner sa caméra. Durant le suivi, *l'élève* suit le *professeur* du regard. Le changement de la direction du *professeur* induit automatiquement un changement de la direction de regard de *l'élève*. Le mouvement du corps ne change de direction que lorsque la tête se stabilise dans une autre direction. Les détails de l'implémentation neuronale de ce mécanisme sont expliqués dans [5]. Ce mécanisme permet une segmentation précise du mouvement perçu et un bon suivi de la cible.

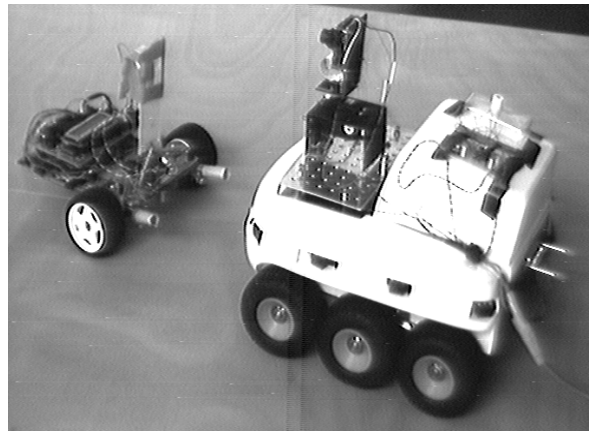


Figure 3: Situation typique de suivi entre 2 robots.

4. Apprentissage d'une séquence

L'architecture PerAc utilisée est présentée fig. 4. L'apprentissage d'une séquence motrice induite par le réflexe de suivi (le lien entre MI et MO) est réalisé par le bloc de prédiction (PO).

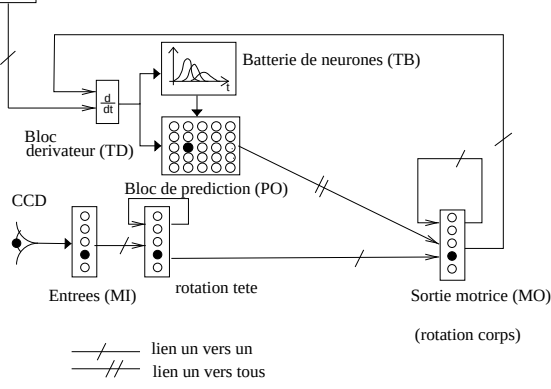


Figure 4: Schéma général de l'architecture PerAc utilisé. CCD - caméra CCD, MI - Entrée du mouvement, MO - Sortie motrice, TD - Dérivateur temporel, TB - Batterie de neurones, PO - Sortie de prédiction.

Les mouvements de *l'élève* sont caractérisés par des changements de son orientation. Les changements d'orientation sont caractérisés par une transition ON/OFF des neurones du bloc MO. Cette transition est détectée par le bloc dérivateur (TD), qui calcule une dérivée temporelle des entrées provenant de MO. L'activation d'un neurone dans TD est modulée par l'activation d'une motivation.

L'utilisation d'une batterie de neurones (TB) avec des constantes de temps différentes (fig. 5), permet au groupe PO une décomposition du signal provenant de TD et l'apprentissage de l'intervalle de temps entre 2 motifs différents de TD (très voisine d'une transformée en ondellette). L'activation d'une cellule TB est décrite par eq. 1.

$$Act_{j,l}^{TB}(t) = \frac{1}{m_j} \cdot \exp - \frac{(t - m_j)^2}{2 \cdot \sigma_j} \quad (1)$$

ou j est l'index de la cellule dans la batterie, l est l'index du neurone TD et de la batterie associée, m_j et σ_j les constantes de temps associées.

L'activation d'un neurone TD met à zéro l'activité de la batterie TB correspondante et permet la modification des poids entre TB et PO (eq. 2).

$$W_{po(i,j)}^{tb(j,l)} = \begin{cases} \frac{Act_{j,l}^{TB}}{\sum_{j,l} (Act_{j,l}^{TB})^2} & \text{si } Act_j^{TD} \neq 0 \\ \text{non modifié} & \text{sinon} \end{cases} \quad (2)$$

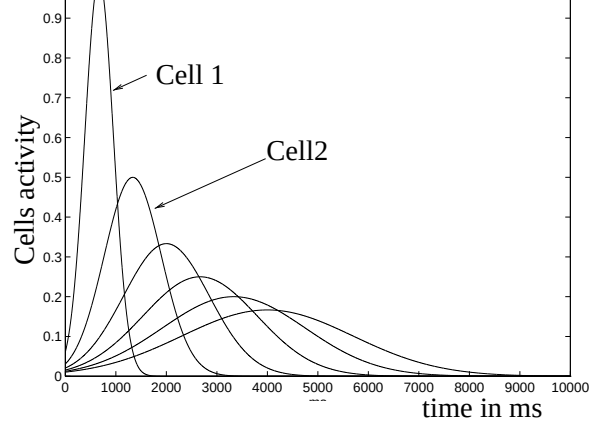


Figure 5: L'activité temporelle d'une batterie de neurones TB.

Une cellule PO s'active seulement quand sont potentiel (eq. 3) atteint son maximum. L'activation d'une cellule PO correspond à la prédiction d'un événement (fig. 6).

$$Pot_{i,j}^{PO} = \sum_l W_{po(i,j)}^{tb(j,l)} \cdot Act_{j,l}^{TB} \quad (3)$$

Pour bien illustrer le comportement de notre réseau, nous présentons à l'entrée du groupe TD la séquence d'événements "1", "2", "3", "1", "2", "3", ..., (fig. 6a). A la première présentation de l'événement 2, le neurone connecté au neurone codant "2" dans TD et à la batterie correspondant au "1" apprend à prédire "2" à partir de "1". Les fois suivantes l'activation de la prédiction de "2" se fait juste avant que "2" s'active (apprentissage de la durée de l'intervalle, fig. 6b).

Pour apprendre une séquence, *l'élève* doit détecter et apprendre les changements d'orientation de son corps. L'apprentissage de chaque séquence est modulé par une variable interne qu'on peut lier à une motivation ou à une émotion. Au début de chaque comportement *l'élève* considère qu'il est dans l'orientation 0 degrés et il refait la séquence en fonction de son orientation de départ.

5. Applications

Nous avons testé notre architecture dans des expériences en temps réel avec des robots KOALA. Le robot *élève* réussit à apprendre

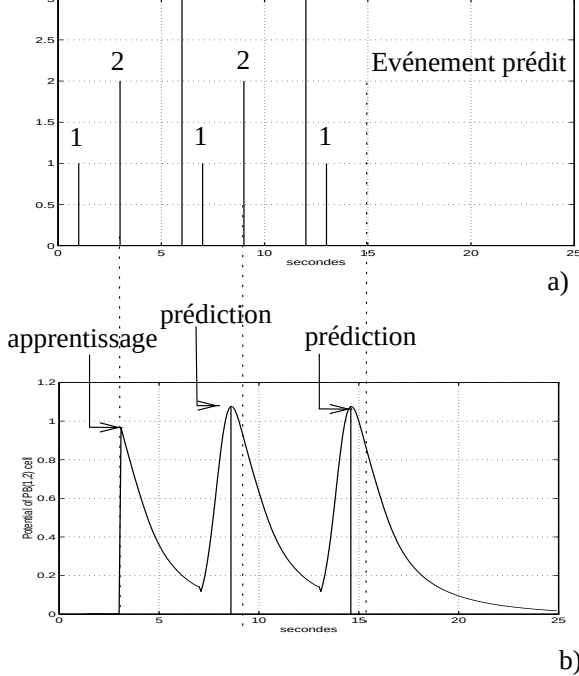


Figure 6: a) Présentation d'une séquence. a) Le potentiel et l'activation du neurone qui prédit l'événement 2 après l'événement 1. Le neurone s'active quand son potentiel atteint le maximum.

plusieurs trajectoires (voir fig. 7) et à reproduire leur séquençage temporel de manière correcte. Les différences entre la trajectoire du professeur et celle de l'élève étant de moins de 10 % en orientation.

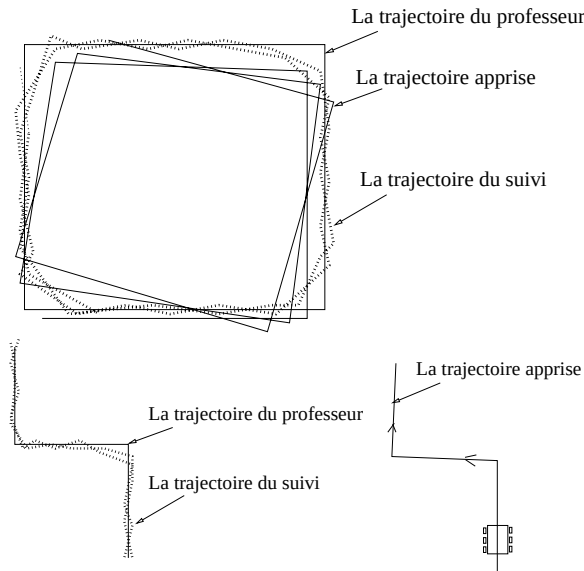


Figure 7: 2 exemples de trajectoires apprises.

Le choix de la trajectoire exécutée est modulé par une variable interne qui peut

expériences, cette variable modulatrice a été directement liée à certains senseurs en infra-rouge du robot.

Notre travail montre la possibilité d'utiliser l'imitation pour l'apprentissage de comportements qu'il aurait été impossible de faire apprendre par essai et erreur. De plus, l'utilisation de l'architecture PerAc permet la réutilisation des outils déjà développés dans notre laboratoire pour la navigation, la prédiction, etc ... Dans un avenir proche, nous allons essayer d'intégrer des mécanismes d'apprentissage de discrimination visuelle pour que notre robot apprenne à suivre la cible même en l'absence de mouvement ou pour qu'il puisse éviter de suivre des objets inintéressants.

References

- [1] P. Bakker and Y. Kuniyoshi. Robot see, robot do : An overview of robot imitation. In *AISB Workshop on Learning in Robots and Animals*, Brighton, UK, 1996.
- [2] J.P. Banquet, J.L. Contreras-Vidal, P. Gaussier, and Y. Burnod. *Fundamentals of neural network modelling for neuropsychologists*, chapter The cortical-hippocampal system as a multirange temporal processor: A neural model. MIT Press, Boston, 1996.
- [3] P. Gaussier, C. Joulain, S. Moga, M. Quoy, and A. Revel. Autonomous robot learning: What can we take for free ? In *International Symposium on Industrial Electronics - ISIE'97*, Guimarães, Portugal, July 1997. IEEE.
- [4] P. Gaussier and S. Zrehen. Moving the frontiers between robotics and biology, special issue, edited by p. gaussier. *Robotics and Autonomous Systems*, 16, 1995.
- [5] J.P. Banquet P. Gaussier, S. Moga and M. Quoy. From perception-action loops to imitation processes: A bottom-up approach of learning by imitation. *accepted to Applied Artificial Intelligence*.

N.B. Le lien entre la structure TD, TB, PO et la modélisation biologique de l'hippocampe est immédiat si nous considérons que TD correspond au cortex entorhinal, TB au gyrus dentelé et PO à la région CA3 (voir [2]).