

Planification ou conditionnement : où est le problème ?

A. Revel, P. Gaussier

ETIS ENSEA/UCP, 6, Av. du Ponceau 95014 Cergy CEDEX

email {revel,gaussier}@ensea.fr

1 Introduction

Nous cherchons à créer un mécanisme de contrôle générique qui permettrait à un robot autonome d'apprendre à réaliser plusieurs tâches complexes durant la même "existence" (approche animat [7]) en utilisant une caméra CCD comme capteur principal. Dans cette optique, nous pensons que s'inspirer de la psychologie et de la neurobiologie peut nous donner de bonnes pistes pour construire des architectures "adaptées" au traitement des flux sensori-moteurs (SM) et utiliser "gratuitement" les régularités de l'environnement sans restreindre les capacités d'adaptation du système.

On distingue en psychologie cognitive deux théories majeures. Le comportementalisme a été introduit par Pavlov dans les années 30 [8] et suggère que la majeure partie des capacités d'apprentissage s'expliquent par des mécanismes de conditionnement. Le cognitivisme, quant à lui, a été introduit par Tolman [9]. Cette théorie ne nie pas les capacités de conditionnement, mais suggère que des comportements complexes tels que ceux de découverte de solution de labyrinthes ne peuvent s'expliquer que si l'on considère que les animaux disposent de "cartes cognitives" de l'environnement et qu'ils les utilisent pour planifier leurs actions.

L'intelligence artificielle a bien sûr tenté de donner écho à ces résultats de psychologie. La première approche a été de limiter "l'intelligence" à sa partie "cognitive". De nombreux systèmes de manipulation de symboles ont ainsi été développés pour permettre la planification des "actions". Cependant des difficultés sont apparues lorsqu'il a fallu con-

fronter ces systèmes à la réalité. En fait, Har-nad a mis en évidence la difficulté intrinsèque de ces systèmes à s' "ancrer" au monde réel (symbol grounding problem - [6]).

Du fait de ces écueils, une nouvelle approche de l'intelligence artificielle est apparue. L'idée principale défendue dans cette approche est que l'intelligence "émerge" de l'interaction de différents niveaux de boucles perception-action construites directement à partir des capteurs et des effecteurs. La première étape a été de réaliser des comportements "réflexes" en connectant directement les capteurs aux effecteurs sans autres traitements intermédiaires (phototaxie ou évitement d'obstacles — [2]). De manière à ajouter une certaine flexibilité, il a été nécessaire d'introduire des mécanismes d'apprentissage. De ce point de vue, le paradigme comportementaliste a servi d'inspiration pour la construction de ces règles. L'étape suivante serait de doter ces systèmes de capacités de planification.

Dès que les situations rencontrées présentent suffisamment de régularités, planifier est une bonne stratégie. En effet, une carte cognitive est un outil efficace pour se souvenir comment sortir d'un labyrinthe. En revanche, la planification est inopérante dès que l'environnement devient imprédictible (murs coulissants, changement de labyrinthe). Le conditionnement peut quant à lui être appris si l'on suppose que les intersections sont marquées par des motifs visuels particuliers indiquant où est la sortie. Dans le contexte animat, le conditionnement instrumental est un bon moyen de faire apprendre un comporte-

information frustrée sur son succès.

Dans un problème de labyrinthe, la difficulté consiste à résoudre le problème d'associations alors que le signal de renforcement ne concerne pas une seule action mais la série d'actions qui a amené l'animat au but (cf. fig. 1).

De manière à résoudre les difficultés engendrées par les problèmes dus au renforcement retardé, nous avons proposé une règle d'apprentissage neuronale qui modélise les capacités de test d'hypothèses (Probabilistic Conditioning Rule — PCR [5]). La résolution du problème du labyrinthe a ainsi pu être testée en pratique sur notre robot Koala en implémentant l'architecture de contrôle PerAc (architecture générique pour l'apprentissage d'associations sensori-motrices — voir [4]).

Ce type de règle de conditionnement nous permet donc d'élaborer des architectures de contrôle pour l'apprentissage de tâches relativement complexes, mais nous devons noter qu'à aucun moment le robot ne "choisit" le comportement qu'il doit effectuer. Dans le cas où une même situation serait associée à deux actions, il est donc nécessaire d'introduire un mécanisme permettant de choisir une action par rapport aux différentes motivations du système.

2 Planification motivée

Dans les années 30, Tolman mis au point une expérience qui remettait en cause l'hypothèse comportementaliste. L'expérience consistait à introduire 3 groupes de rats dans un labyrinthe : le premier groupe était un groupe récompensé tout le temps, le deuxième un groupe récompensé uniquement au bout de 11 jours. Selon l'hypothèse comportementaliste, les rats du premier groupe auraient du avoir des résultats bien supérieurs à ceux du deuxième groupe. En fait, les résultats montrèrent qu'après le onzième jour, les résultats du premier et deuxième groupe étaient équivalents. Tolman en conclut que, bien qu'ils ne fussent pas récompensés avant le onzième jour ("apprentissage latent"), les rats du deuxième groupe avaient réussi à se construire une

cognitive").

Pour construire une telle représentation, on peut utiliser une carte de neurones entièrement inter-connectés. Lorsqu'un neurone s'active, il reste actif pendant un certain temps et si un deuxième neurone s'active, alors un simple apprentissage hebbien peut permettre d'apprendre le lien synaptique entre ces deux neurones.

En généralisant ce mécanisme à toutes les situations rencontrées, le graphe des relations topologiques entre les situations peut être appris ainsi que le lien avec un neurone codant la motivation.

Durant la planification, l'information apprise lors de l'exploration peut être utilisée pour atteindre le but lié à une motivation donnée. La solution adoptée consiste à rétro-propager l'activation de la motivation pour que l'activité des neurones de la carte soit le reflet de leur distance au but (en nombre de transitions). La règle d'activation proposée est la suivante :

$$y_i = \max_j (W_{ij} \cdot y_j)$$

L'activité des neurones étant bornée par 1, les poids des connexions doivent être bornés par une valeur $W_{Max} < 1$. Il faut ensuite itérer jusqu'à la stabilisation des valeurs des activités. Cet algorithme est équivalent à l'algorithme de Bellman de découverte du plus court chemin dans un graphe [1].

Ce mécanisme de planification est d'autant plus intéressant qu'il permet, de plus, de proposer un plan qui fait un bon compromis entre l'intensité des motivations qui forcent le système à atteindre certains buts et la longueur des chemins pour les atteindre.

3 Lier niveau SM et planification

Pour que la planification puisse s'appliquer, une première idée serait de combiner sur les mêmes neurones les mécanismes d'apprentissage sensori-moteur et de planification. Malheureusement, un neurone unique ne peut décider quelle situation est reconnue et propager l'information motivationnelle

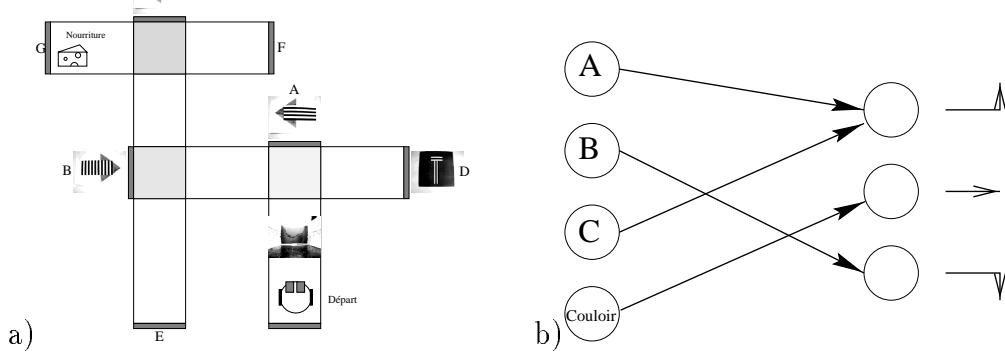


Figure 1. a) Le labyrinthe utilisé et les motifs perçus correspondant (obtenus à partir de la caméra du robot). b) Les associations devant être faites pour atteindre le but.

de but en sous-but. Il y aurait en effet ambiguïté sur la raison de son activation.

Deux niveaux au moins sont donc nécessaires de manière à séparer les 2 flux d'information : l'un correspondant au niveau "but", l'autre à la reconnaissance des situations et à leur association à un mouvement (niveau "sensori-moteur"). Une situation est en faite représentée par deux neurones liés réciproquement, l'un se trouvant au niveau but, l'autre au niveau sensori-moteur. Pour l'apprentissage, l'information doit "remonter" du niveau reconnaissance au niveau but de manière à construire la carte cognitive. A l'inverse, pour la planification, l'information provenant du niveau but doit "redescendre" de manière à sélectionner l'action la plus appropriée vis à vis de la motivation courante. Ce mode de fonctionnement correspond à celui des colonnes corticales modélisées par Burnod [3].

Une solution serait alors de superposer directement le niveau planification décrit dans la section précédente au niveau sensori-moteur.

Dans l'exemple figure 2, la situation A est associée au mouvement "avancer", B est associé à la fois à "tourner à droite" et à "tourner à gauche", C à la satisfaction de la motivation et D est un cul-de-sac.

Durant la planification, le système doit pouvoir déclencher la séquence d'actions qui permet d'atteindre C à partir de A, sachant que le but est en C. Or, lorsque l'animat ar-

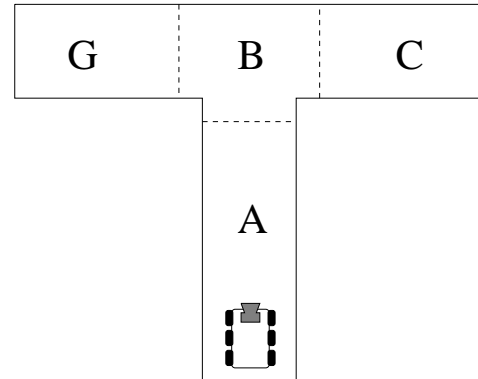


Figure 2. Exemple de labyrinthe pour la planification

rive en B il n'y a aucun moyen de choisir quelle action doit être effectuée puisque B est lié à 2 mouvements !

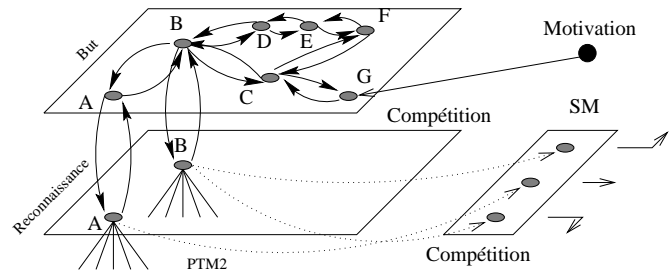


Figure 3. Il est impossible de planifier en utilisant uniquement la reconnaissance des situations car elles peuvent être liées à plusieurs mouvements.

En fait, il devrait y avoir deux représentations distinctes associées chacune à un mouvement. Une solution consiste à utiliser une représentation composite correspondant à la transition d'une situation vers une autre. Par exemple, la transition temporelle de A vers B peut être apprise par un neurone codant

effectuer pour se rendre de A à B .

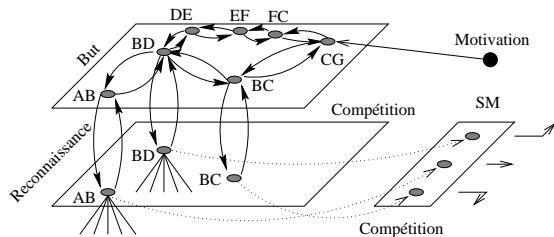


Figure 4. Utilisations des transitions pour la planification.

Travailler à partir des transitions permet donc de planifier sans ambiguïté les mouvements à effectuer pour atteindre un but (voir fig. 5). La résolution de ce problème a été très largement aidée par les travaux de modélisation neurobiologique de l'hippocampe que nous faisons par ailleurs. A l'inverse, ces travaux de modélisation neurobiologique ont profité des réflexions systémiques impliquées dans l'approche robotique.

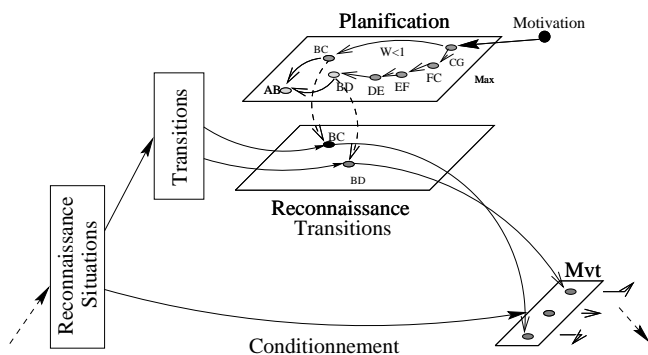


Figure 5. Elaboration des transitions dans un modèle de l'hippocampe.

4 Conclusion

Nous avons proposé qu'un système de planification devait s'appuyer sur un mécanisme d'apprentissage d'associations sensori-motrices pour pouvoir permettre l'intégration et le contrôle des différents flux d'information nécessaires à la planification. Par ailleurs, nous avons mis en exergue l'importance de la dynamique dans l'élaboration de représentations internes et des liens réciproques entre le niveau de planification et le niveau sensori-moteur.

dans une même approche, le contrôle robotique et l'intelligence artificielle. Nous proposons pour cela une architecture qui permet à la fois l'apprentissage de conditionnement et la planification en tenant compte que les aspects motivationnels ne sont pas seulement utilisés au niveau planification, mais sont sans doute en cause dans l'élaboration des signaux de renforcement utilisés pour conditionner un animal.

Bibliographie

- [1] R.E. Bellman. On a routing problem. *Quarterly of Applied Mathematics*, (16):87–90, 1958.
- [2] V. Braitenberg. *Vehicles : Experiments in Synthetic Psychology*. Bradford Books, Cambridge, 1984.
- [3] Y. Burnod. *An adaptive neural network: the cerebral cortex*. Masson, 1989.
- [4] P. Gaussier, C. Joulain, A. Revel, and J.P. Banquet. Are shaping techniques the correct answer for the control of an autonomous robot? In *UKACC International Conference on Control'96*, University of Exeter, September 1996. IEE.
- [5] P. Gaussier, A. Revel, C. Joulain, and S. Zrehen. Living in a partially structured environment: How to bypass the limitation of classical reinforcement techniques. *Robotics and Autonomous Systems*, 20:225–250, 1997.
- [6] S. Harnad. The symbol grounding problem. *Physica D*, 42:335–346, 1990.
- [7] J.A. Meyer and S.W. Wilson. From animals to animats. In MIT Press, editor, *First International Conference on Simulation of Adaptive Behavior*. Bradford Books, 1991.
- [8] I.P. Pavlov. *Conditioned Reflexes*. Oxford University Press, 1927.
- [9] E.C. Tolman. Cognitive maps in rats and men. *The Psychological Review*, 55(4), 1948.