Un algorithme évolutionniste pour l'auto-apprentissage de groupes de robots mobiles autonomes

Philippe Lucidarme et Alain Liégeois
lucidarm@lirmm.fr, liegeois@lirmm.fr
LIRMM (UMR 55060)
161, rue Ada, 34392 Montpellier cedex

Jean-Louis Vercher et Reinoud Bootsma
Vercher@laps.univ-mrs.fr
UMR 6152 « Mouvement et Perception »
CP 910, avenue de Luminy, Marseille cedex 09

Résumé
Cette publication présente un algorithme d'apprentissage automatique de la fonction d'exploration d'un ensemble de robots mobiles. Il est basé sur les principes des algorithmes génétiques mais fonctionne sans le superviseur qui donne habituellement une note à la performance de chaque individu et effectue un classement de la population selon ce critère. Ici, chaque individu estime sa propre performance. Une autre différence avec les algorithmes classiques est également liée à l'absence de superviseur : les opérations de croisement ne peuvent se faire que lorsque deux individus se rencontrent physiquement. Les résultats de simulations sur ordinateur montrent en particulier le temps de convergence de l'apprentissage de la population en fonction du nombre de robots. Le bon fonctionnement a été également vérifié expérimentalement en utilisant plusieurs minirobots réactifs. Les principes sont appliqués à une étude de modélisation de chaînes sensori-motrices chez l'Homme. Les premiers résultats d'une tâche de ralliement de cible sont présentés.

1. Introduction

Habituellement, les tâches complexes qui nécessitent la coopération de plusieurs robots mobiles, pour l'agriculture, l'exploration des planètes, ou les applications industrielles et domestiques, sont pratiquement programmées à l'avance. De plus, elles mettent en jeu un niveau très cognitif : connaissance d'une carte de l'environnement, capacité à se localiser et à localiser les autres agents, algorithmes de planification de trajectoires et de mouvements, etc. (Pinchard, Liégeois et Emmanuel 1995 ; Balch et Arkin 1995 ; Bothelho et Alami 1999) Seuls des travaux sur l'émergence d'un comportement collectif intelligent à partir d'individus à capacités limitées traitent une forme d'apprentissage automatique de coopération (Mataric 1997 ; Parker 1998 ; Simonin et Ferber, 2000). En vue de nous rapprocher du comportement naturel d'apprentissage chez les êtres vivants, il nous est apparu naturel d'étudier la possibilité d'utiliser les principes de l'évolutionnisme, notamment ceux des algorithmes génétiques (Goldberg 1989) qui transforment une population d'individus en de nouveaux plus performants. Quelques travaux de recherche ont déjà abordé cette voie (Floreano et Mondana 1994 ;Lin, Xiao et Michaelevicz 1994 ; Koza et Rice 1991 ; Floreano et Urzelei 2000). D. Floreano conclut que la procédure ne converge pas vers une préférence franche pour une règle particulière parmi celles inspirées des Neurosciences. Pour cela, nous recherchons une méthode qui permet à chaque robot de construire par lui-même ses règles de renforcement à partir de l'auto-évaluation de l'efficacité de son comportement sensori-moteur, sous la forme d'une autosatisfaction compatible avec l'architecture proposée par (Simonin et Ferber 2000).

2. L'algorithme d'apprentissage

2.1 Description de la tâche collective

Il s'agit, pour un ensemble de robots mobiles à deux roues motrices et équipés d'émetteurs et de récepteurs infrarouges, d'apprendre à explorer toute la surface d'un univers physique de manière sûre, c'est-à-dire en évitant les collisions avec les murs, les obstacles et les autres robots. Actuellement, on suppose une population homogène, constituée d'individus identiques, ayant les mêmes capacités de perception, de mobilisation, de communications et de calcul embarqué

2.2 Description et codage chromosomique d'un individu

Les entrées du système sensori-moteur sont délivrées par un ensemble d'émetteurs et de récepteurs infrarouges entourant le robot. Ils correspondent à la vision grossière et quasi-

instantanée, dite basse fréquence, chez les être vivants. Elles correspondent aux 5 états donnés sur le Tableau 1.

Etat 1	Pas d'obstacle
Etat 2	Obstacle à gauche
Etat 3	Obstacle à droite
Etat 4	Obstacle en face
Etat 5	Blocage

Tableau 1. Les différents états du système

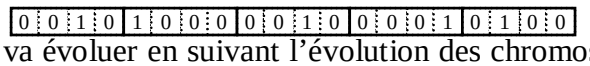
Comportement 1	Avancer tout droit
Comportement 2	Tourner à droite
Comportement 3	Tourner à gauche
Comportement 4	Reculer

Tableau 2. L'ensemble des réactions

A l'autre extrémité, la sortie, du système de contrôle-commande, l'individu a les comportements élémentaires donnés sur le Tableau 2.

Le chromosome d'un robot est la concaténation (une chaîne) de N mots dans un sous-espace de l'espace binaire $\{0,1\}^M$. N est le nombre d'entrées, et M le nombre de sorties. Bien entendu ici, un seul bit vaut 1 dans chaque mot. Un exemple est donné sur la Figure 1.

Figure 1. Un exemple de chaîne chromosomique



La population de robots va évoluer en suivant l'évolution des chromosomes sous l'action des opérateurs génétiques.

2.3 L'algorithme d'apprentissage

Initialisation. Au début, les chaînes sont remplies de mots choisis aléatoirement.
L'indice de performance individuel. Comme nous recherchons ici un apprentissage non supervisé, chaque robot doit s'auto-évaluer, mais il n'est pas conscient de la performance collective. Chaque individu numéro *i* calcule sa performance d'après :

$$R_{N(i)}(i) = (1-\alpha(i))R_{N(i-1)}(i) + \alpha(i)F_{N(i)}(i) \qquad (1)$$

où

N(i) est le nombre de périodes élémentaires de la chaîne sensori-motrice depuis le début de l'auto-évaluation,
$R_{N(i)}(i)$ est la récompense estimée à cette date,
$F_{N(i)}(i)$ est la récompense instantanée à cette date.

Ici, la récompense instantanée est la distance parcourue par unité de temps sensori-moteur sans reculer ni être bloqué. Ainsi, la récompense intégrée est la distance moyenne parcourue depuis l'initialisation ou depuis le dernier croisement ou la dernière mutation.
Croisement. La plupart des chercheurs supposent une communication globale entre les individus, ce qui est parfois impossible et de toute façon prend du temps et nécessite un protocole compliqué et inutile puisque seulement deux individus peuvent être sélectionnés en même temps. Nous avons préféré une communication simple et locale. Ainsi, seuls deux robots qui se rencontrent se communiquent leur code génétique et leur indice de performance, et procèdent au croisement à condition qu'un temps suffisamment long se soit écoulé depuis la dernière mutation ou le dernier croisement pour chacun, afin que les valeurs de performances soient fiables. On démontre qu'après plusieurs croisements, la performance moyenne n'a pas diminué. Toutefois les mutations sont indispensables pour garantir la convergence vers un optimum absolu.
Mutation. La stratégie suivante est adoptée : le robot n'a pas muté récemment et sa performance est faible. Nous faisons varier la probabilité de mutation en fonction de la performance : des fonctions sigmoïdes et en escalier sont essayées.

3. Résultats des simulations

La distance de communication est celle de nos robots mobiles expérimentaux. L'intervalle minimum de temps entre deux mutations ou croisements est de l'ordre de cinquante pas élémentaires, ce qui laisse le temps à deux robots qui viennent de se croiser de s'éloigner suffisamment pour ne pas recommencer juste après.
La Figure 2 montre le résultat d'une exploration (la surface en blanc) à la fin d'un apprentissage (tous les robots ont la bonne association perception-action). Les petits cercles

noirs sont les robots, et les gros sont les obstacles. Nous avons étudié la durée de l'apprentissage en fonction du nombre de robots. Elle décroît jusqu'à 10 robots puis ne change plus. On n'a donc pas intérêt à multiplier le nombre d'individus.

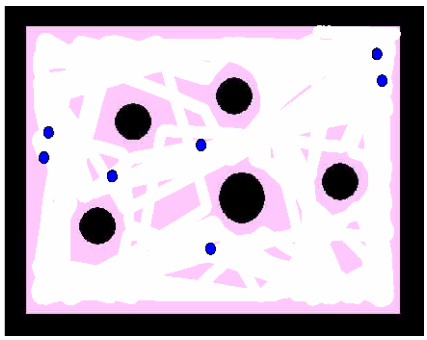


Figure 2. Surface explorée à la fin d'un apprentissage.

Figure 3. Exemple d'expérimentation.

4. Expérimentations avec des robots réels

Les robots mobiles utilisés ont été conçus et réalisés au LIRMM (P. Lucidarme, P. Rongier et A. Liégeois 2001). Leur forme est un cylindre d'environ 12 cm de diamètre et 13 cm de haut. 16 émetteurs et 8 récepteurs infrarouges sont régulièrement espacés sur la périphérie. Le cycle sensori-moteur élémentaire dure 15 ms. La portée des communications infrarouges peut être ajustée. Elle est typiquement de 0,5 m. La vitesse maximum a été limitée à 30 cm/s par sécurité pour le matériel. La durée minimum entre deux croisements est de 3 secondes. Un PC embarqué (80486 DX avec horloge à 66 MHz) assure le traitement de l'information. La figure 3 montre un exemple d'expérimentation avec 4 robots. On retrouve les résultats prévus par les simulations. La convergence est obtenue après seulement quelques minutes.

5. Extrapolation à une tâche individuelle de ralliement de cible

Ce type d'expérience a été réalisé simultanément avec des humains au laboratoire Mouvement et Perception, et au LIRMM avec un des robots qui a appris l'évitement d'obstacle. La procédure d'évaluation consiste à comparer la trajectoire des robots après apprentissage, à celle d'individus humains se déplaçant vers un but tout en évitant des obstacles. Pour les robots , la cible est une balise émettrice infrarouge, à une fréquence différente de celle de la détection d'obstacles, qui déclenche une attirance chez le robot dès qu'il la voit, d'autant plus forte que le signal perçu est de plus grande amplitude. La fonction de renforcement (l'autosatisfaction) est ainsi modifiée. La Figure 4 montre une des trajectoires de simulations, comparée à celle d'un opérateur humain dans une tâche de ralliement de cible avec évitement d'obstacle.

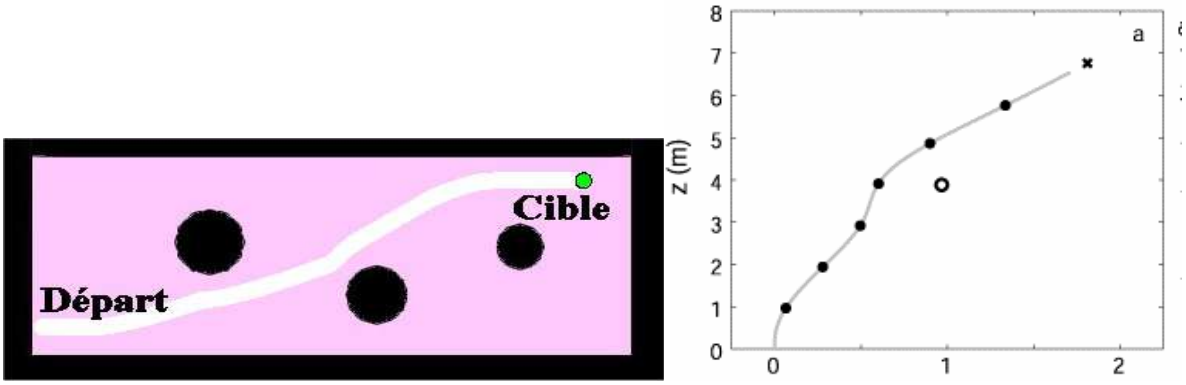


Figure 4. Exemple de ralliement de cible, à gauche avec le robot (travaux du LIRMM), à droite chez l'homme (d'après Warren et coll. 2000).

6. Conclusion et perspectives

Nous avons présenté un algorithme génétique à convergence sûre pour l'apprentissage des chaînes sensori-motrices de robots mobiles réactifs. Le renforcement est une fonction simple d'autosatisfaction (P. Lucidarme, O. Simonin et A. Liégeois 2002). Les premières comparaisons avec des expériences menées chez l'homme, pour des comportements sensori-moteurs non conscients nous encouragent à poursuivre cette voie de modélisation. Des résultats plus complets seront présentés au cours des Journées NSI'2002. Par le passé, les Neurosciences ont souvent puisé dans les Sciences Physiques des modèles théoriques et des approches expérimentales. A leur tour, on constate que les Sciences Physiques vont chercher dans les Sciences du Vivant des modèles permettant d'envisager la réalisation d'artefacts dont la conception est d'inspiration biologique et permettra à ceux-ci de s'adapter seuls à des environnements inconnus. La robotique apporte en retour la possibilité d'implémenter et de tester des hypothèses de contrôle. Nous envisageons par exemple d'appliquer les algorithmes développés au LIRMM au contrôle d'un manipulateur mobile et de comparer la cinématique obtenue ainsi à celle produite par un sujet humain dirigeant sa main vers une cible visuelle. Les robustesses des processus d'apprentissage et d'adaptation à des perturbations sera analysée respectivement chez l'Homme et pour la machine, et comparées qualitativement et quantitativement.

Références

T. Balch and R. Arkin, 1995, Motor schema-based formation control of multiagent robot teams, Int. Conf. On Multiagent Systems, p. 10-16.

S. Botelho et R. Alami, 1999, Multirobot Cooperation in the Martha Project, ICRA'99, p. 1234-1239.

D. Floreano and F. Mondada, 1994, Automatic Creation of an Autonomous Agent : Genetic Evolution of a Neural-Network Driven Mobile Robot, SAB-3, Brighton, p. 421-430.

D. Floreano and J. Urzelai, 2000, Evolutionary Robots with On-line Self-Organization, Neural Networks, Vol . 13, p. 397-404.

D.E. Goldberg, 1989, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley.

J.R. Koza and J.P. Rice, 1991, Genetic Generation of both the Weignts and Architecture for a Neural Network, IJCNN-91, Seattle, Vol. 2, p. 397-404.

H.S. Lin, X. Xiao and Z. Michalewicz, 1994, Evolutionary Navigator for a Mobile Robot, ICRA'94, San Diego, Vol. 3, p. 2199-2204.

P. Lucidarme, P. Rongier et A. Liégeois, Implementation and evaluation of a Reactive Multi-Robot System, AIM'01, Como, p. 165-170.

P. Lucidarme, O. Simonin et A. Liégeois, 2002, Implementation and evaluation of a Satisfaction/Altruis-Based Architecture for Multi-Robot Systems, ICRA'02, Washington, May 2002.

M. Mataric, 1997, Behavior-based Control Examples from Navigation, Learning, and Group Behavior, J. of Experimental and Theoretical A. I., 9(2-3), p. 323-336.

L.E. Parker, 1998, ALLIANCE : An ARCHITECTURE for Fault-tolerant Multirobot Cooperation, IEEE Trans on Robotics and Automation, Vol. 14, N° 2, p. 220-240.

O. Pinchard, A. Liégeois and T. Emmanuel, 1995, A Genetic Algorithm for Outdoor Robot Path Planning, 4th Int. Conf. On Intelligent Robotic Systems, Karsruhe, IOS Press, p. 413-419.

O. Simonin and J. Ferber, 2000, Modeling Self Satisfaction and Altruism to Handle Action and Reactive Cooperation, SAB'00 Proceedings Supplement, p. 314-323.

WH Jr Warren, BA Kay, WD Zosh, AP Duchon, S Sahuc., 2000, Optic flow is used to control human walking, Nat Neurosci.;4(2):213-6.