BOSAR : Bibliothèque d'Outils Simulink pour l'Apprentissage par Renforcement

G. Laurent, E. Piat, C. Adda, N. Le Fort-Piat

Laboratoire d'Automatique de Besançon - UMR CNRS 6596
24 rue Alain Savary, F-25000 Besançon
glaurent@ens2m.fr, epiat@ens2m.fr, cadda@ens2m.fr, npiat@ens2m.fr

1. Contexte des recherches sur l'apprentissage par renforcement au LAB

1.1. *Microrobotique et micro-assemblage*

Le Laboratoire d'Automatique de Besançon (LAB) développe depuis une dizaine d'années une activité de recherche importante dans le domaine de la microrobotique et des microsystèmes de production. L'activité en microrobotique a pour objectif général de concevoir, réaliser et commander des systèmes micromécatroniques complexes, en particulier des microrobots et des cellules microrobotiques de production, destinés à manipuler des objets de dimensions caractéristiques typiquement comprises entre 1 mm et 1 μm. Les microrobots ne sont pas forcément des dispositifs de tailles micrométriques mais simplement de petite taille (quelques dizaines de centimètre au plus). Ils ont par contre des effecteurs de taille micrométrique afin de manipuler des objets également micrométriques.

L'activité concernant les cellules microrobotiques de production se place à une échelle conceptuelle plus large et porte sur l'étude et la réalisation de systèmes de micro-assemblage capables de produire en petites ou moyennes séries, des équipements pour le marché futur de la micromécatronique, des microsystèmes et du secteur biomédical[1].

1. L'enjeu est ici de développer des moyens de production extrêmement flexibles qui soient capables de se reconfigurer en fonction du produit à fabriquer. Par ailleurs, ces moyens de production, appelés communément *micro-usines*, doivent être à « l'image » des produits assemblés,

L'apprentissage par renforcement en tant que méthode pour concevoir des contrôleurs capables de commander correctement des microrobots ou des cellules microrobotique a été introduit au LAB en 2000.

1.2. *Contexte spécifique à la microrobotique*

D'un point de vue expérimental, on constate pendant les tâches de micromanipulation que de nombreux paramètres liés au micromonde sont souvent mal maîtrisés (rugosité des surfaces, humidité, distribution de charges électrostatiques, etc.) et induisent des microforces (forces Van Der Waals, forces électrostatiques etc.) non contrôlées. Le processus continu à commander prend alors un caractère aléatoire marqué qui nuit à la répétabilité des tâches à effectuer. En l'absence de modèle fiable, la synthèse de contrôleurs par les approches traditionnelles de l'Automatique est donc difficile.

Dans ce contexte spécifique, nous nous sommes intéressés à la commande échantillonnée bloquée (i.e. discrétisée) de processus continus (car physiques) à caractère aléatoire marqué. On peut noter que l'apprentissage par renforcement dans sa formulation classique se prête bien à la synthèse d'une commande optimale d'un processus discrétisé et de nature stochastique. L'apprentissage par renforcement a de ce fait une légitimité évidente dans le contexte de la microrobotique. Par contre, comme nous sommes tributaire d'un apprentissage en temps réel, la taille et la décomposition de l'espace d'état et de l'espace de commande devient un paramètre critique pour obtenir des apprentissages dans des temps « raisonnables ». Les temps d'apprentissage visés sont ici de l'ordre de l'heure à quelques dizaines d'heures pour des périodes d'échantillonnage de quelques millisecondes à quelques dizaines de millisecondes (une commande est générée à chaque période d'échantillonnage du système).

1.3. *Axes de recherche développés en apprentissage par renforcement*

Les axes de recherche développés en AR au LAB visent à diminuer le temps d'apprentissage d'une commande optimale sur les microrobots du laboratoire. Par conséquent, nous nous intéressons uniquement à l'apprentissage « temps réel » sur des systèmes physiques existants en microrobotique et généralement insuffisamment ou non modélisés. Trois directions sont actuellement en cours d'investigation :

– la décomposition adaptative des espaces d'état et de commande pendant l'apprentissage afin de diminuer la cardinalité de l'espace état-commande. Cette réflexion est tout juste amorcée.

---

c'est-à-dire avoir un encombrement réduit (de l'ordre du mètre) et une faible consommation énergétique. La vision communément admise est alors d'avoir des unités de micro-assemblage « pluggable » sur une plate-forme véhiculant l'information (issues des capteurs et à destination des effecteurs), l'énergie ainsi que certains flux de matière (gaz, liquide ...).

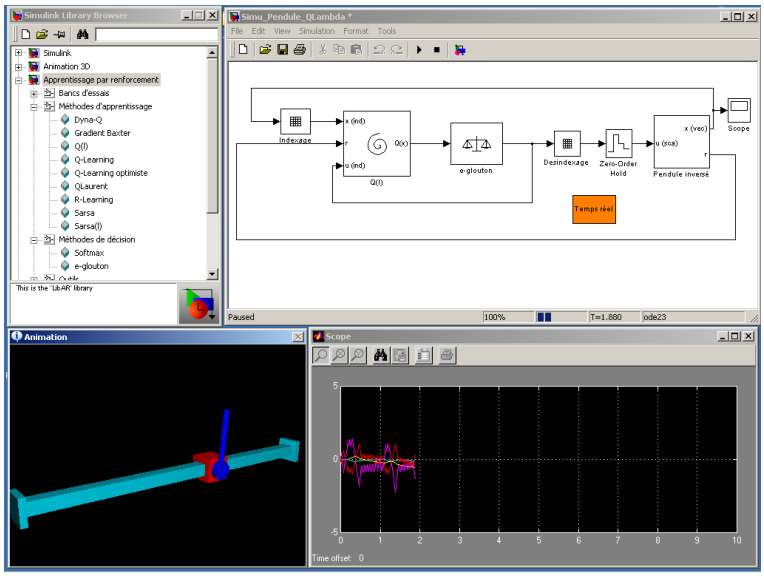

Figure 1. *Capture d'écran montrant BOSAR en fonctionnement*

– les méthodes indirectes (model-based) qui permettent de « sortir » de la boucle de commande l'optimisation des fonctions de valeurs (traitement hors ligne).

– les approches multi-agents qui permettent une décomposition judicieuse de l'espace de commande dans certains cas de figure.

2. BOSAR : Bibliothèque d'Outils Simulink pour l'Apprentissage par Renforcement

2.1. *Objectifs*

Afin de pouvoir développer rapidement des algorithmes d'apprentissage par renforcement pour commander des systèmes dynamiques classiques (continus ou discrets), nous avons développé un ensemble d'outils qui fonctionnent avec le logiciel de calcul numérique Matlab-Simulink. Cette bibliothèque de blocs Simulink baptisée BOSAR permet ainsi une étude et une évaluation aisée des algorithmes d'apprentissage par renforcement. BOSAR couvre les domaines suivants :

– apprentissage par renforcement direct ou indirect,

– contrôle de systèmes continus ou discrets,

– contrôle de systèmes complètement ou partiellement observables.

Ses principaux atouts sont :

– l'utilisation de la convivialité de Simulink pour la création, le test de nouvelles simulations et la comparaison avec des algorithmes existants,

– l'utilisation des fonctions de Matlab pour le post-traitement des données,

– l'utilisation du solveur de Simulink pour la simulation de systèmes continus (intégration à pas variable).

2.2. *Contenu*

Le contenu de BOSAR s'organise en quatre rubriques :

– la rubrique « méthodes d'apprentissage » contient les algorithmes d'apprentissage (Q-Learning, Q(λ), Sarsa(λ), Dyna-Q, R-Learning, OLPOMDP, ...),

– la rubrique « méthodes de décision » contient les méthodes de sélection d'action (ϵ-glouton, softmax, ...),

– la rubrique « bancs d'essais » contient les systèmes académiques de tests (grid-world discret, pendule inversé simple, pendule inversé sur chariot, mountain car, puck world, ...),

– la rubrique « outils » contient des blocs facilitant la mise en place de certaines simulations.

2.3. *Perspectives*

BOSAR se veut libre (licence GPL) et ouverte à la communauté d'apprentissage par renforcement. L'ambition est d'obtenir un outil performant contenant les principaux algorithmes de la littérature actuelle. Une version de BOSAR sera mise en ligne d'ici aux journées de PDMIA.

3. Domaines d'application présents et futurs

Les algorithmes développés au LAB ont été utilisés pour commander deux dispositifs physiques :

– un mini-pousseur de cellules (thèse de G. Laurent en 2002) baptisé WIMS[2]-MACRO. Ce dispositif est une réplique à l'échelle 20 d'un micro-pousseur de cellules de 10 à 100 μm de diamètre qui a été développé au LAB. Les deux systèmes utilisent un mode d'actionnement magnétique pour faire évoluer le pousseur sur un plan. Le WIMS-MACRO est capable, soit d'atteindre une succession de points de consigne en position tout en évitant les objets présents, soit d'aller chercher et de pousser des objets dans une direction donnée.

2. WIreless Micromanipulation System

– un micro-positionneur d'objets micrométriques ou millimétriques baptisé METI[3]. Ce dispositif est constitué de 3 pointes ayant chacune un degré de liberté en translation et actionnées par des translateurs de précision micrométrique. Chaque pointe peut pousser en un point donné le micro-objet à positionner. Ce dispositif sert actuellement de banc d'essais pour les approches multi-agents et les méthodes indirectes.

Les perspectives d'application futures au LAB se situent principalement dans le cadre des microsystèmes de production qui constituent une activité de recherche en croissance forte. Des applications en émergence sont notamment :

– le positionnement optimal de micro-imageurs qui observent la réalisation d'une ou plusieurs tâches dans le micromonde,

– la commande de processus de micro-assemblage ayant un caractère aléatoire (aléas du micromonde) afin de garantir au mieux la réalisation de chaque tâche,

– l'optimisation de la configuration des micro-usines soit à l'intérieur d'un poste de travail, soit entre différents postes.

3. Micro ÉTau Intelligent