Apprentissage de la verticalisation sur un humain virtuel

Renaud Durlin, Camille Salaun et Olivier Sigaud
Université Pierre et Marie Curie - Paris 6
4 Place Jussieu, 75252 Paris Cedex 05, France
durlin@poleia.lip6.fr, Camille.Salaun@robot.jussieu.fr, Olivier.Sigaud@lip6.fr

Pour commander un humanoïde, nous nous intéressons aux modèles computationnels d'apprentissage tirés des sciences du mouvement. L'enjeu est de développer une activité scientifique combinant un objectif appliqué – obtenir par apprentissage une loi de commande pour un humanoïde – et un objectif fondamental – participer via la modélisation robotique à l'étude de l'apprentissage du mouvement humain. Dans cette contribution, nous exposons des résultats préliminaires obtenus en simulation avec un modèle inspiré de MMRL (Doya et al., 2002) sur une tâche de verticalisation.

Dispositif expérimental

Nous utilisons le simulateur HuMAnS (Wieber et al., 2006). L'état de l'humain virtuel est décrit par l'ensemble des positions angulaires des segments. Le répertoire d'actions consiste à appliquer séparément un couple positif ou négatif sur chacune des 3 articulations principales de l'humain dans le plan (chevilles, genoux et bassin), en imposant un couple identique pour les articulations droites et gauches. La tâche consiste à partir accroupi et à terminer debout. La fonction d'évaluation $r(x_t)$ récompense les états dans lesquels la hauteur de la tête atteint une valeur telle que l'humain virtuel est debout, et à punir les états dans lesquels elle descend en dessous d'une valeur seuil inférieure à son altitude de départ.

Apprentissage de la commande

Nous utilisons un algorithme d'apprentissage par renforcement indirect appliqué dans un espace continu. Comme MMRL, notre modèle apprend un modèle direct de la dynamique de l'humain simulé et un modèle de la fonction de valeur associée à la tâche de verticalisation.

L'objectif de l'apprentissage du modèle direct est de trouver la fonction f qui donne l'état suivant estimé $\hat{x}_{t+1}$ en fonction de l'état courant x_t et de la commande u_t. Pour apprendre f, nous utilisons l'algorithme IMTI (Potts,), après avoir constaté l'inefficacité des perceptrons multi-couches.

L'objectif de l'apprentissage de la fonction de valeur $V(x_t)$ est d'associer à chaque état une valeur telle que, si le système cherche à chaque instant à atteindre l'état de plus grande valeur parmi les états atteignables à l'instant suivant, alors il résout sa tâche de manière optimale. Cette fois, nous utilisons un perceptron multi-couches avec une règle de *TD-learning*. Le choix de l'action est régi par l'équation : $u_{t+1} = \arg\max_u [r(f(x_t, u_t)) + V(f(x_t, u_t))]$, avec 2 % de choix aléatoire pour garantir une forme d'exploration.

Expériences et résultats

Nous testons notre apprentissage de la commande soit isolément, soit en utilisant la loi de commande produite pour affiner une loi de commande écrite manuellement, qui ne réussit pas tout à fait à réaliser la tâche.

Afin d'évaluer la difficulté de la tâche, nous commençons par déterminer le nombre de succès obtenus avec un choix aléatoire de l'action. Sur 10.000 essais, ce nombre est nul avec notre loi de commande seule, et de YY % en complément de la loi de commande manuelle.

Nous utilisons ensuite notre algorithme d'apprentissage en réalisant 100 séries de 100 essais et en mesurant le taux de séries qui convergent vers un succès systématique. Ce taux reste nul avec notre loi de commande seule et de Y % en complément de la loi de commande manuelle. Le nombre de succés sur les 10.000 essais correspondants est de YYY.

Discussion et travaux futurs

Nos résultats en l'absence de la loi de commande manuelle montrent que résoudre une tâche de verticalisation suppose de disposer d'un répertoire d'actions finement ajusté.

Quand cette loi de commande manuelle existe, l'apprentissage permet de l'affiner significativement, même si notre modèle est encore loin de converger systématiquement.

Pour améliorer le taux de convergence, plutôt que de nous reposer pour l'apprentissage de la fonction de valeur sur des perceptrons multi-couches dont nous avons mesuré l'insuffisance, nous envisageons d'avoir recours au formalisme de la commande optimale, d'abord au travers des *Linear Quadratic Controllers*, comme c'est déjà le cas dans MMRL, puis en nous tournant vers le modèle d'Emmanuel Guigon (Guigon et al., 2006), ce qui nous rapprochera de l'étude du mouvement humain.

Références

Doya, K., Samejima, K., Katagiri, K., & Kawato, M. (2002). Multiple model-based reinforcement learning. *Neural Computation*, *14*, 1347–1369.

Guigon, E., Baraduc, P., & Desmurget, M. (2006). Computational motor control : Redundancy and invariance. *J. Neurophysiology*, *in press*.

Potts, D. Incremental learning of linear model trees. *Proceedings ICML'04* (pp. 663–670).

Wieber, P.-B., Billet, F., Boissieux, L., & Pissard-Gibollet, R. (2006). The HuMAnS toolbox, a homogeneous framework for motion capture, analysis and simulation. *Proceedings IWHHFR*.