Mehdi Khamassi
Chercheur CNRS
ISIR - Sorbonne Université (ex UPMC)
Pyramide – Tour 55 – Boîte courrier 173
4 place Jussieu, 75005 Paris, France
e-mail : mehdi.khamassi@upmc.fr

Sujet de stage de Master ouvert à candidature (2019) :

Rôle des replay off-line dans la coordination de systèmes d'apprentissage par renforcement pour la navigation

Contexte

L'équipe AMAC de l'ISIR a récemment développé différents algorithmes d'apprentissage par renforcement dotés de mécanismes de replay off-line – permettant de rejouer mentalement certains éléments de la mémoire épisodique pour accélérer l'apprentissage – et les a comparé en simulation sur une tâche de navigation dans un multiple-T-maze (Gupta et al., 2010) pour étudier quels algorithmes permettent de mieux rendre compte des dynamiques de réactivation des cellules de lieux de l'hippocampe chez le rat pendant des périodes de pause en éveil ou des périodes de sommeil (Cazé et al., 2018). Nous avons en particulier étudié comment différentes méthodes de prioritisation de l'information en mémoire épisodique permettaient ou non de reproduire les proportions observées expérimentalement de réactivation de cellules de lieux dans le même ordre que celui observé pendant l'éveil (forward), dans l'ordre inverse (backward), dans un désordre apparent (random), qui peut tout de même cacher une certaine structure quand il est généré par certains types d'algorithme suivant une règle précise. Un constat important de ce travail est que des simulations en environnement discret (monde à case) ou en continu ne donnent pas les mêmes effets d'une proportion donnée de replay sur la performance d'apprentissage, et ne permettent donc pas de reproduire de la même façon les courbes de performance des rats.

Par ailleurs, l'équipe a aussi testé un certain nombre d'algorithmes d'apprentissage par renforcement sans replay sur des expériences de navigation sur robots à roue (Caluwaerts et al., 2012) ainsi que d'autres expériences sur robots humanoïdes simulés (Renaudo et al., 2014, 2015a,b). Ces expériences ont eu jusqu'à maintenant pour objectif principal d'étudier les propriétés de différents mécanismes de coordination entre différents types d'apprentissage par renforcement dits *model-based (MB)* (apprenant un modèle du monde pour la planification délibérée de l'action vers un but) et *model-free (MF)* pour l'apprentissage d'association stimulus-réponse, en partant d'un modèle de neurosciences computationnelles pour cette coordination permettant d'expliquer un certain nombre de données expérimentales chez le rat (Dollé et al., 2018, 2008). Ces expériences ont de plus permis de montrer une meilleure performance robotique par une combinaison d'apprentissages MB et MF que par une combinaison de multiples apprentissages MF (Khamassi et al., 2006).

Objectif

L'objectif du stage est d'étudier l'effet de différents algorithmes de replay off-line sur les performances d'apprentissage en situation de navigation sur un robot à roue turtlebot. L'objectif secondaire est d'étudier dans quelle mesure l'utilisation d'un mécanisme de replay off-line change la façon dont les systèmes d'apprentissage MB et MF doivent être coordonnés, permettant par exemple un transfert de connaissances du MB au MF par simulation mentale de séquences d'actions dans le modèle pendant les phases off-line. Ceci doit permettre in fine la mise au point d'un nouveau modèle computationnel d'apprentissage par renforcement avec replay dont la robustesse ait été établi par l'expérimentation sur robot réel, pour ensuite faire un retour vers la biologie et étudier dans quelle mesure ce modèle permet de rendre compte d'un certain nombre de données expérimentales sur la réactivation off-line des cellules de lieux de l'hippocampe pendant des tâches de navigation chez le rat.

Méthodes

Les expériences menées pendant ce stage se feront sur le labyrinthe de navigation robotique actuellement en place à l'ISIR. Ce labyrinthe constitue une adaptation du labyrinthe de Tolman (Tolman, 1948), incluant plusieurs chemins de longueur différente pour atteindre un but. Il y a de plus des phases de changements abruptes de la tâche, avec soit l'introduction d'un obstacle bloquant un des chemins, ou le changement de la position du but. Les expériences menées actuellement étudiant un algorithme d'apprentissage combiné MB et MF mais sans mécanisme de replay off-line, le travail de stage consistera à intégrer à cet algorithme et comparer les différentes méthodes de replay étudiées dans (Cazé et al., 2018). Il s'agira dans un premier temps de comparer leurs performances sur robot avec celles obtenues en simulation. Dans un deuxième temps, il s'agira d'étudier si l'ajout de mécanismes de replay change la dynamique d'adaptation du robot aux changements de l'environnement. En particulier, le robot passe-t-il moins de temps à ré-utiliser son système d'apprentissage MB après un changement de l'environnement lorsqu'il est doté de mécanismes de replay permettant un transfer supposé plus rapide d'information entre MB et MF, que dans le cas où le robot n'est pas doté de mécanismes de replay ? Le robot évite-t-il à certains moments d'utiliser son système MB lorsque les replay lui permettent de faire converger plus rapidement son système MF ? Est-ce que les expériences ainsi réalisées permettent de prédire différentes dynamiques d'alternance entre phases de contrôle MB et MF sur le robot par rapport aux expériences précédemment réalisées avec un modèle computationnel sans replay ? Enfin, si le temps le permet, une troisième phase du stage pourrait consister à revenir vers la simulation de protocoles neurobiologiques plus récents de navigation pour voir si le nouveau modèle ainsi conçu permet de mieux rendre compte de données expérimentales chez le rat.

Mots-clefs :

neurosciences computationnelles, prise de décision, navigation, apprentissage par renforcement, replay, hippocampe, robotique cognitive.

Institut des Systèmes Intelligents et de Robotique
Sorbonne Université - CNRS / UMR 7222
Pyramide Tour 55 - Boîte courrier 173
4 Place Jussieu - 75252 PARIS cedex 05 - France
Tél. +33 (0)1 44 27 51 31 / 51 41 - Fax +33 (0)1 44 27 51 45

Références

Caluwaerts, K., Staffa, M., N'Guyen, S., Grand, C., Dollé, L., Favre-Félix, A., Girard, B., Khamassi, M., 2012. A biologically inspired meta-control navigation system for the psikharpax rat robot. Bioinspiration & Biomimetics 7 (2), 025009.

Cazé, R., Khamassi, M., Aubin, L., Girard, B., 2018. Hippocampal replays under the scrutiny of reinforcement learning models. Journal of neurophysiology 120 (6), 2877–2896.

Dollé, L., Chavarriaga, R., Guillot, A., Khamassi, M., 2018. Interactions of spatial strategies producing generalization gradient and blocking : A computational approach. PLoS computational biology 14 (4), e1006092.

Dollé, L., Khamassi, M., Girard, B., Guillot, A., Chavarriaga, R., 2008. Analyzing interactions between navigation strategies using a computational model of action selection. In : International Conference on Spatial Cognition. Springer, pp. 71–86.

Gupta, A. S., van der Meer, M. A. A., Touretzky, D. S., Redish, A. D., 2010. Hippocampal Replay Is Not a Simple Function of Experience. Neuron 65 (5), 695–705.

Khamassi, M., Martinet, L.-E., Guillot, A., 2006. Combining self-organizing maps with mixtures of experts : application to an actor-critic model of reinforcement learning in the basal ganglia. In : International Conference on Simulation of Adaptive Behavior. Springer, pp. 394–405.

Renaudo, E., Devin, S., Girard, B., Chatila, R., Alami, R., Khamassi, M., Clodic, A., 2015a. Learning to interact with humans using goal-directed and habitual behaviors. In : Ro-Man 2015, Workshop on Learning for Human-Robot Collaboration.

Renaudo, E., Girard, B., Chatila, R., Khamassi, M., 2014. Design of a control architecture for habit learning in robots. In : Conference on Biomimetic and Biohybrid Systems. Springer, pp. 249–260.

Renaudo, E., Girard, B., Chatila, R., Khamassi, M., 2015b. Respective advantages and disadvantages of model-based and model-free reinforcement learning in a robotics neuro-inspired cognitive architecture. Procedia Computer Science 71, 178–184.

Tolman, E. C., 1948. Cognitive maps in rats and men. Psychological review 55 (4), 189–208.

Institut des Systèmes Intelligents et de Robotique
Sorbonne Université - CNRS / UMR 7222
Pyramide Tour 55 - Boîte courrier 173
4 Place Jussieu - 75252 PARIS cedex 05 - France
Tél. +33 (0)1 44 27 51 31 / 51 41 - Fax +33 (0)1 44 27 51 45