

Proposition de Stage Recherche Master

*Comment coder la récompense
dans des systèmes d'apprentissage par renforcement?*

Proposé par:

Thomas BORAUD, Institut des Maladies Neurodégénératives (IMN) et
Frédéric ALEXANDRE, IMN et Inria

LaBRI, équipe CombAlgo

frederic.alexandre@labri.fr 06 30 38 71 94

Ce projet peut être financé à 100% par l'IMN

L'apprentissage par renforcement fait l'objet d'études approfondies aussi bien dans le domaine du Machine Learning (ex: algorithme de Q-Value, Sarsa) que dans celui de la modélisation bio-inspirée (ex: algorithme PVLV). Il est notable que ces deux approches se rejoignent jusqu'à un certain point, en mettant en avant l'erreur de prédiction de récompense (supposée être codée dans les systèmes bio-inspirés par l'activité dopaminergique phasique) comme critère central des règles d'apprentissage.

Ces deux approches se rejoignent également par le peu d'importance qu'elles donnent à une notion qui peut pourtant paraître centrale: le codage de la récompense. Deux éléments sont en particulier à souligner. D'une part, le niveau tonique de la dopamine semble avoir également un rôle important dans le choix motivé des réponses. D'autre part, parler d'une récompense est quelque peu réducteur dans la mesure où elle peut être transmise par des canaux bien différents et les structures réputées coder cette récompense comptant des centaines de milliers de neurones, on peut également supposer une représentation quelque peu élaborée. Sur ces deux points, ces modèles classiques restent relativement muets.

Ce constat est une première motivation pour se poser la question, au niveau de la modélisation, du codage de la récompense dans les systèmes motivés. Une deuxième raison sera l'interprétation de données physiologiques acquises à Bordeaux par des équipes de biologistes partenaires et posant

également une question originale et liée à ces considérations. En effet, la plupart des modèles informatiques ou biologiques de l'apprentissage par renforcement considèrent le cas où la récompense est donnée... comme récompense dans le cas d'une réponse comportementale adaptée à une certaine tâche. Les biologistes proposent de considérer un cas où la même quantité de récompense est donnée mais de façon non contingente aux réponses comportementales. On observe alors une dégradation de l'apprentissage au niveau comportemental et les biologistes se proposent ici d'aller mesurer l'activité des structures dopaminergiques pour comprendre les changements neuronaux associés.

En résumé, ce stage se propose d'étudier ces différents éléments pour en abstraire la proposition bio-inspirée d'un codage plausible de la récompense. En premier lieu, les algorithmes classiques bio-inspirés ou pas seront passés en revue. Ensuite, les différentes propriétés de la récompense dont de tels systèmes sont supposés rendre compte seront énumérées et décrites précisément. Sur cette base, on pourra alors proposer (partant éventuellement d'un algorithme existant) un codage et une dynamique intégrant au mieux ces différents éléments. Des tâches classiques d'apprentissage par renforcement ainsi que les principaux résultats de l'expérimentation mentionnée pourront être confrontés à ce modèle.