

# Sujet 3

## RL événementiel sur Animat

Proposé par : Alain Dutech, Hervé Frezza-Buet

Equipe : Biscuit

### Informations générales

Encadrants    Alain Dutech, Hervé Frezza-Buet  
Adresse        LORIA, Campus Scientifique - BP 239, 54506 Vandœuvre-lès-Nancy  
Téléphone     03 83 59 20 95  
Email          alain.dutech@loria.fr, herve.frezza-buet@centralesupelec.fr  
Bureau        C 044

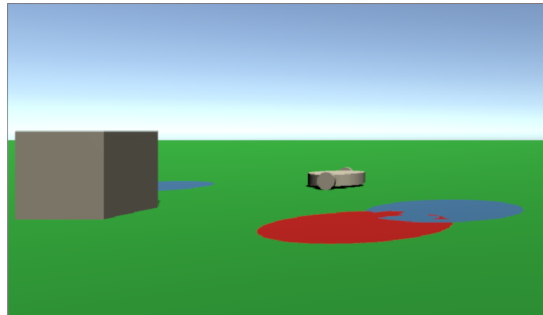


FIGURE 2 – Vue de l’environnement artificiel de l’Animat. Unity3D.

### Motivations

Dans cette étude, nous nous intéressons à l’apprentissage du comportement d’un agent doté d’une physiologie artificielle. Il s’agit d’une approche animat [1] où l’agent doit apprendre à se comporter de sorte à maintenir ses constantes vitales dans un domaine de viabilité (homéostat). L’animat est un système physique évoluant à temps continu selon les lois de l’environnement. Dans cette étude, nous pourrions considérer à la fois des systèmes réels (plateformes robotiques mobiles comme les turtlebots ou les quadricoptères parrot) ou un système simulé, développé au sein de l’équipe BISCUIT, et reposant sur le moteur 3D Unity.

Ce problème d’apprendre un comportement (une séquence d’actions) pour maximiser un critère numérique (fonction d’une récompense liée à l’homéostat) est adressé classiquement dans le cadre formel de l’apprentissage par renforcement[2]. Ce formalisme repose sur des interactions synchrones entre l’agent et l’environnement ( $t$ ,  $t+1$ , ...). Cette discrétisation temporelle n’est pas adaptée dans le cas d’interactions continues avec l’environnement. Afin de concilier le caractère discret en temps des méthodes de renforcement avec le caractère continu du déroulement du temps avec un système physique, la solution que nous souhaitons explorer est d’identifier des événements dans ce système physique. La succession d’événements revêt alors le caractère discret en temps compatible avec les approches d’apprentissage par renforcement, en définissant un temps « logique » plutôt qu’un temps physique.

### Sujet

Dans un premier temps de l’étude, il s’agira d’implémenter un contrôle cablé à la main pour s’appropriier la plateforme et la dynamique du problème. Cette phase permettra d’identifier la génération d’événements nécessaires au contrôle. Dans un second temps, on s’intéressera à appliquer les méthodes issues de l’apprentissage par renforcement (SARSA, Q-Learning, Acteur/Critique) sur la base du séquençage réalisé par les événements. Cette étude expérimentale permettra d’exhiber la complexité du contrôle d’un système dynamique par des approches séquentielles et d’envisager de nouvelles solutions algorithmiques.

- ROS : <https://wiki.ros.org>
- Unity3D : <https://unity3d.com>

## Cadre du travail

Le stage s'effectuera au Loria, un poste de travail sera fourni. Des notions en apprentissage par renforcement sont attendues. Les travaux seront réalisés en python ou C++ avec ROS<sup>2</sup>. Par ailleurs, ils pourront être menés sur des plateformes réelles ou une simulation développée par l'équipe avec Unity3D.

## Références

- [1] Meyer, J. and Wilson, S. W. *Animat Scholarpedia*, vol 4 (5), doi :10.4249/scholarpedia.1533, 2009.
- [2] R. Sutton and A. Barto, *Reinforcement Learning : An introduction*, 2nd edition, MIT Press, Cambridge, MA, 2018.

---

2. qui pourra être appris à l'occasion