Projet Introduction à l'informatique et à la programmation

Combats de robots

Enseignant : L. Ralaivola

Date : 13 octobre 2017

1 Objectifs du projet

Programmation Python. Ce projet a deux objectifs. En tout premier lieu, il s'agit évidemment pour les étudiantsd de continuer leur familiarisation avec le langage Python au travers d'une mise en pratique sur un sujet ludique, celle des bandits à K-bras. Une interface graphique (en mode texte) sera mise à disposition et, en plus de la simple programmation Python telle que déjà vue en cours, quelques concepts très simples de programmation objet seront discutés.

Bandit à K bras. Un deuxième objectif est de comprendre une problématique scientifique amusante : celle, déjà mentionnée, des bandits à K-bras. Cette problématique mathématique, dont s'est emparée il y a peu de temps la communauté d'apprentissage automatique (qui s'intéresse à faire en sorte que les ordinateurs soient capables d'*apprendre*, comme les êtres humains), trouve son origine dans les machines à sous, autrement appelées bandits manchots [1], que l'on trouve dans les casinos. La question que se pose un joueur qui se rend dans un casino et qui veut jouer sur ces machines, est évidemment de maximiser ses gains. Face à K machines différentes [2], dont chacune rapporte un gain – inconnu du joueur – moyen différent, le joueur est confonté au problème *« exploitation/exploration » :* s'il a essayé M machines différentes, avec $M < K$, et qu'il a identifié en faisant la moyenne de ses gains sur chacune d'elle que la machine m rapporte plus que les autres, il peut décider d'*exploiter* cette connaissance et ne jouer que cette machine à sous. Il se peut néanmoins qu'une autre machine, parmi les $K - M$ non encore essayées, rapporte plus que la machine m : pour le savoir, le joueur est obligé d'essayer ou *explorer* les machines non encore utilisées. Le joueur doit donc définir une stratégie d'exploration/exploitation lui permettant, au bout de T actions de jeu, d'avoir un gain aussi proche que possible qu'un joueur ayant joué T fois sur la machine m^* dont la moyenne des gains associés est la plus élevée. Des stratégies très efficaces existent, dont, en particulier, les stratégies `UCB` (pour Upper Confidence Bound) et ε-`gloutonne`, et il s'agira de les programmer.

2 Combats de robots et créativité

Combats de robots Une illustration de cette problématique de bandits à K bras est la suivante. Des robots d'une même équipe sont présents dans un environnement donné, dans

1. Ces machines d'usage très simple fonctionnent comme suit : il suffit d'y introduire une pièce et d'en actionner le bras pour savoir si/combien on a gagné.
2. Notons que l'on appelle le problème celui des bandits à K bras plutôt que celui des K bandits à 1 bras.

lequel se trouvent des robots adverses. Ces robots ont la possibilité de choisir une parmi K actions possibles à chaque pas de temps : avancer, tourner (dans une des huit directions possibles), rester sur place. Suivant les actions prises, l'environnement renvoie des récompenses plus ou moins élevées aux robots. Pour découvrir la(les) bonne(s) action(s) à choisir, les robots implémentent simplement les stratégies de la littérature pour le problème des bandits à K-bras. Les étudiants doivent donc programmer le comportement de leurs robots grâce à ces algorithmes.

Créativité. Enfin, ce projet doit offrir la possibilité à chacun de laisser libre cours à son imagination et sa créativité. Une extension assez directe du travail consistant à programmer les algorithmes `UCB` et `ε-gloutonne` est celle qui vise à utiliser des versions « contextuelles » des algorithmes de bandits : des éléments de contexte (notamment la position d'autres robots) sont alors pris en compte dans le choix des actions à prendre.