

# 3I005

## TME 2–4 : Projet Exploration / Exploitation

Ariana Carnielli  
Yasmine Ikhelif

### Table des matières

1	Introduction	1
2	Bandits-manchots	2
3	Morpion	2
4	Puissance 4	2
5	Conclusion	2

## 1 Introduction

Ce mini-projet s'intéresse à la problématique de l'exploitation *vs* exploration, qui consiste à choisir, parmi une quantité de ressources limitées, combien de ces ressources vont être utilisées pour explorer le problème, afin d'avoir plus de chances de trouver la meilleure solution possible, et combien seront utilisées pour exploiter cette meilleure solution. Si toutes les ressources sont dépensées pour l'exploration, on aura plus de chances de trouver une solution proche de l'optimale, mais moins de ressources disponibles pour son exploitation. De l'autre côté, arrêter l'exploration trop tôt peut conduire au choix d'une action sous-optimale pendant la phase d'exploitation, ce qui conduit à un gain plus petit à long terme.

Dans ce mini-projet, on illustre la problématique de l'exploitation *vs* exploration dans trois situations. La première, décrite dans la Section 2, s'intéresse à l'exemple des bandits-manchots. Il s'agit de considérer une machine à sous à  $N$  leviers, chacun ayant une probabilité de victoire différente inconnue du joueur. L'objectif est de maximiser le gain, ce qui nécessite un bon équilibre entre l'exploration des différents leviers afin d'avoir une bonne estimation de la probabilité de gain de chacun et l'exploitation du meilleur levier. L'algorithme principal permettant cet équilibre est l'algorithme UCB, dont les détails sont donnés dans la Section 2.

La deuxième situation, décrite dans la Section 3, considérée est celle du jeu de morpion, où trois stratégies de jeu sont implémentées. La première consiste dans une stratégie purement aléatoire, la deuxième est une stratégie de Monte Carlo, qui explore les

actions possibles de façon aléatoire et choisit la meilleure et la dernière, une stratégie de Monte Carlo Tree Search, qui utilise l'algorithme UCB pour optimiser l'exploration des actions possibles.

Finalement, ces mêmes algorithmes, ayant été codés de façon généraliste, sont appliqués à un jeu légèrement plus complexe que morpion, le jeu puissance 4. Les détails de l'implémentation et les résultats obtenus sont donnés dans la Section 4

## **2 Bandits-manchots**

## **3 Morpion**

## **4 Puissance 4**

## **5 Conclusion**

Ce rapport a permis, dans la Section 2, de mettre en évidence la problématique de l'exploitation *vs* exploration à travers l'exemple des bandits-manchots, qui illustre bien l'intérêt d'un algorithme équilibrant exploration et exploitation comme l'algorithme UCB. On a également pu remarquer, dans les Sections 3 et 4, que l'utilisation de cet algorithme dans la stratégie Monte Carlo Tree Search permet d'obtenir des joueurs très performants.