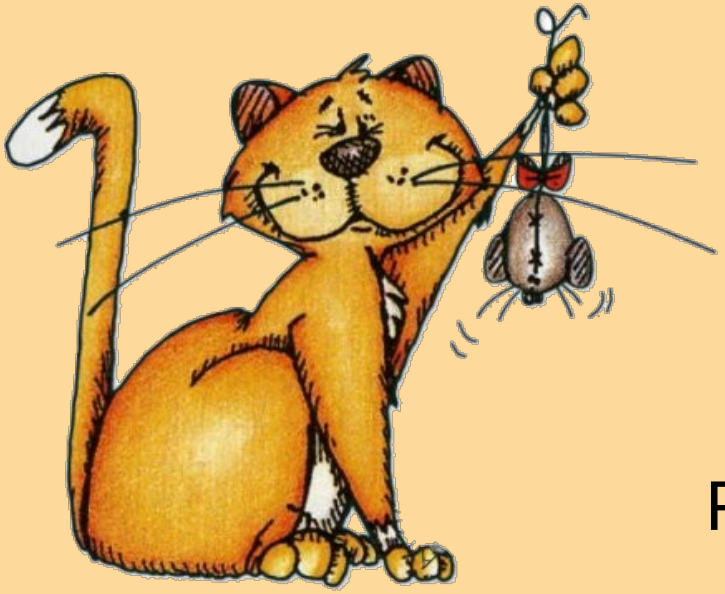
FSA/INGI - 5 septembre 2006

Application du Reinforcement Learning à un jeu de Markov de type évasion-poursuite



Lionel Dricot

Promoteur : Professeur Marco Saerens

Structure

4 composantes à ce mémoire :

1. **Bibliographique :** Reinforcement Learning et jeux de Markov

2. **Théorique :** Analyse du problème et développement d'une méthode de coopération

3. **Technique :** Implémentation d'un framework assez généraliste.

4. **Expérimentale :** Simulations et résultats

1.1 Reinforcement Learning

- Un **agent** au sein d'un **environnement**

- Maximiser le **Reward**
  - Passer un obstacle ?
- Maximiser le **Return attendu**
  - Boucler infiniment sur un reward positif ?
- Maximiser le **Return attendu amorti**
  - paramètre gamma

- Dilemme Exploration - Exploitation

1.2 MDP & Markov Games

- Propriété de Markov : signal qui décrit entièrement un état (jeu d'échec, gravité)

Deterministic Markov Decision Process : à chaque état correspond une valeur (équation de Bellman)

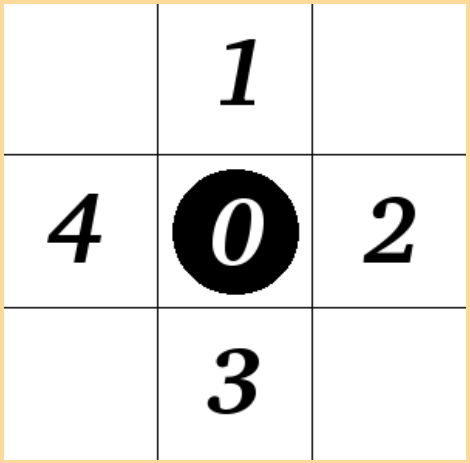
  - Résolution par Q-Learning

$$Q[s,a] = (1-\alpha)Q[s,a] + \alpha(r + \gamma V(s'))$$

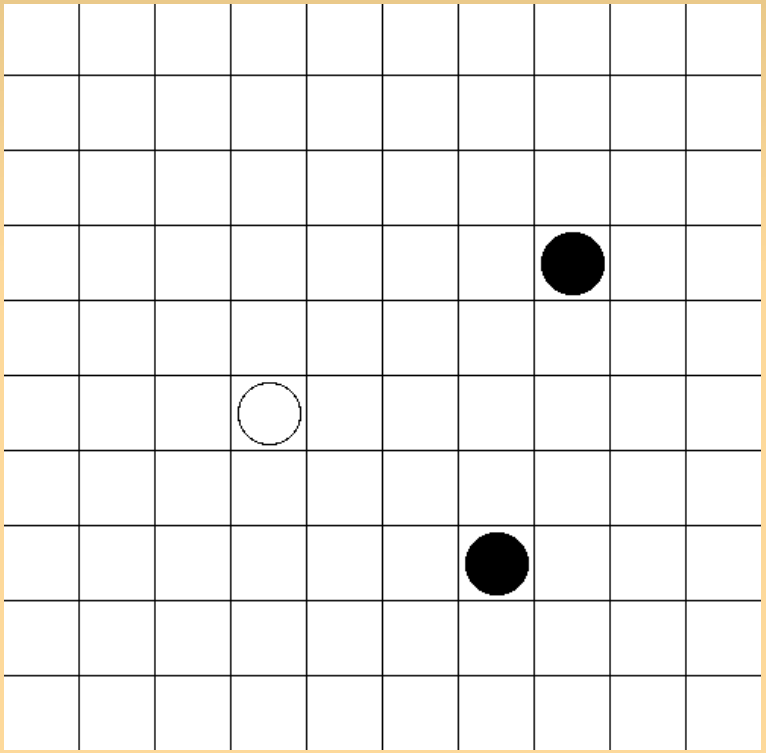
- Jeu de Markov : présence d'agents stochastiques
  - Transposition du Q-Learning aux Markov Games (Littman)

2.1 Le problème chat-souris (1)

- Discret en temps
- Discret spatialement
- Torique
- Alterné
- 5 mouvements possibles par agent



- MDP si souris Best-Escape (BE)
- Markov game si BE Stoch.



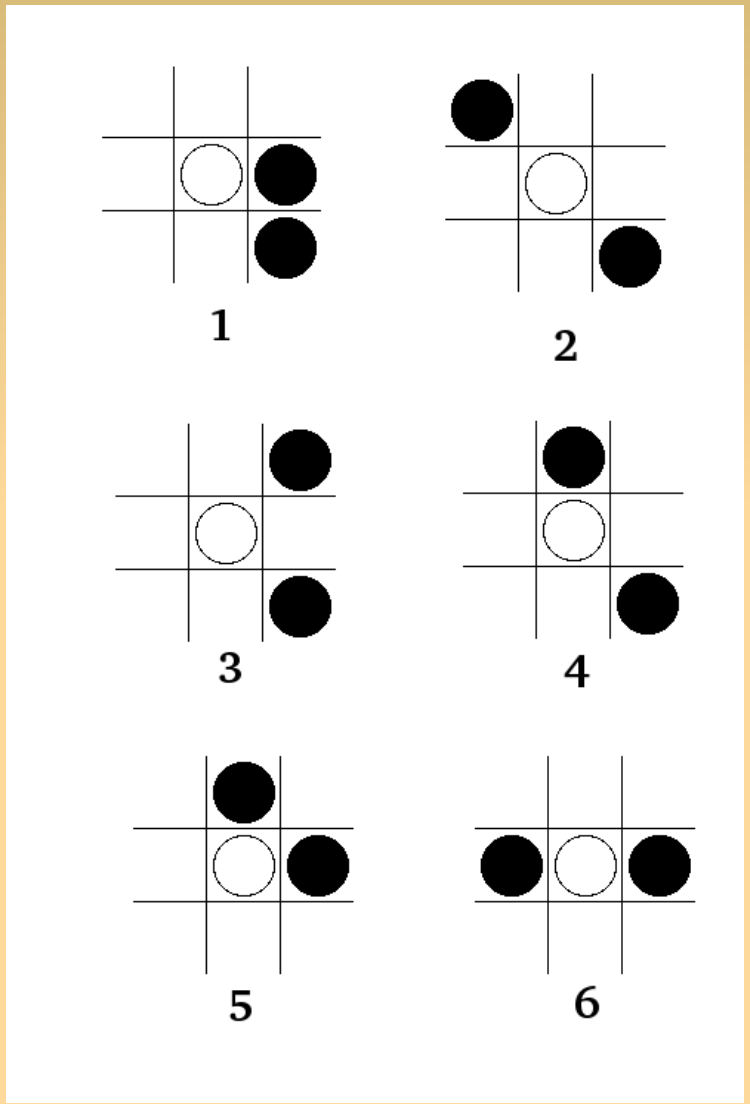
Reward : opposé de la somme des carrés des distances à la souris + extra reward sur la souris

5-6 min

2.1 Le problème chat-souris (2)

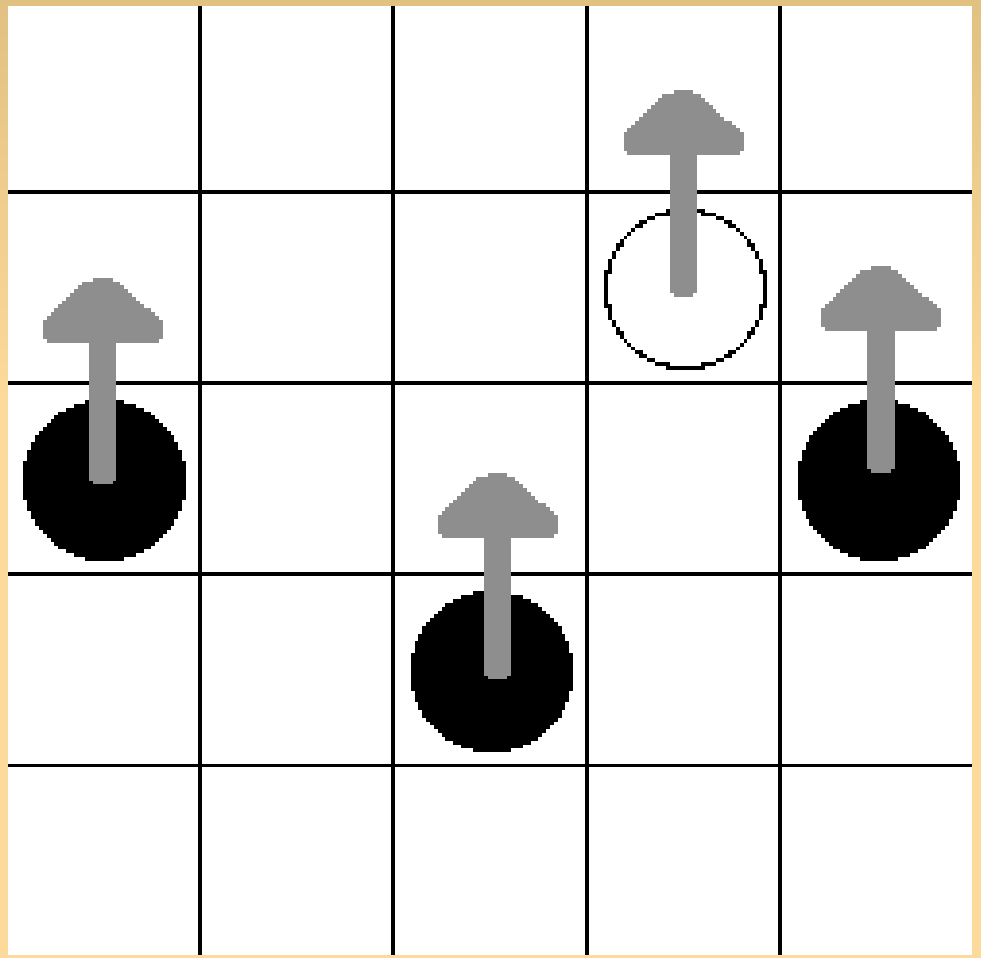
Problème impossible pour 2 chats :

- 3 chats
- fatigue de la souris



2.2 Coopération entre les chats

Problème du parallélisme
->
Coopération entre les chats indispensable !



Les chats sont les **sous-agents** d'un seul et unique **agent**.
L'ensemble des **mouvements** au temps t représente une **action** de l'agent.

2.3 Exploration - Exploitation

3 possibilités de résolution :

1. **Exploration stochastique fixe :** taux constant [Littman, 1994]

2. **Exploration stochastique décroissante :** idem mais avec un taux décroissant (utilisé ici)

3. (perspective) **Entropie et degré d'exploration :** exploration liée à l'état de chaque noeud. Idéal mais plus complexe [Achbany et al, 2006].

3.1 Implémentation

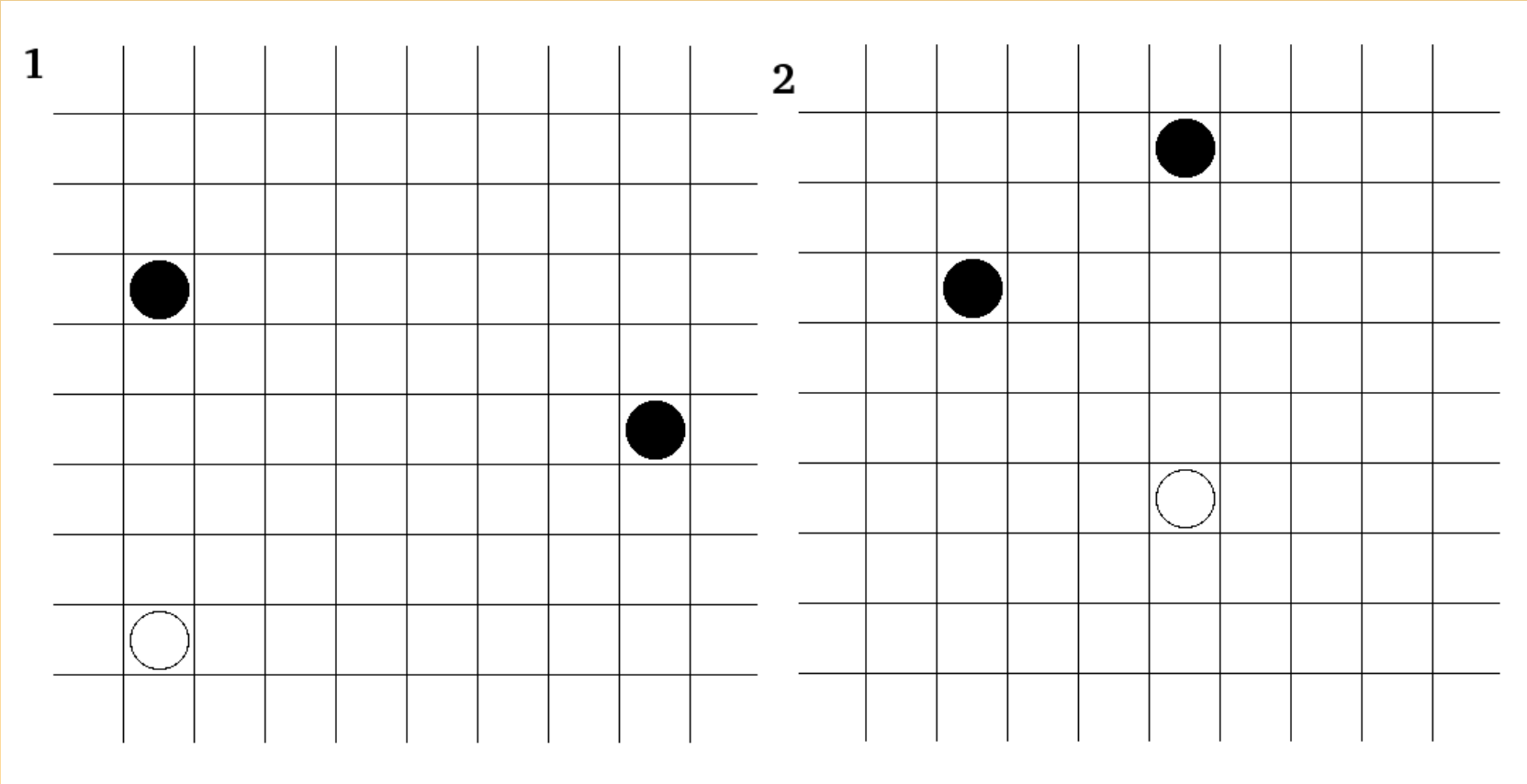
Implémentation d'un framework assez généraliste et modulaire.

Technologies utilisées :

- Python
- LiveWires 2.0 (bibliothèque d'affichage en Tk)

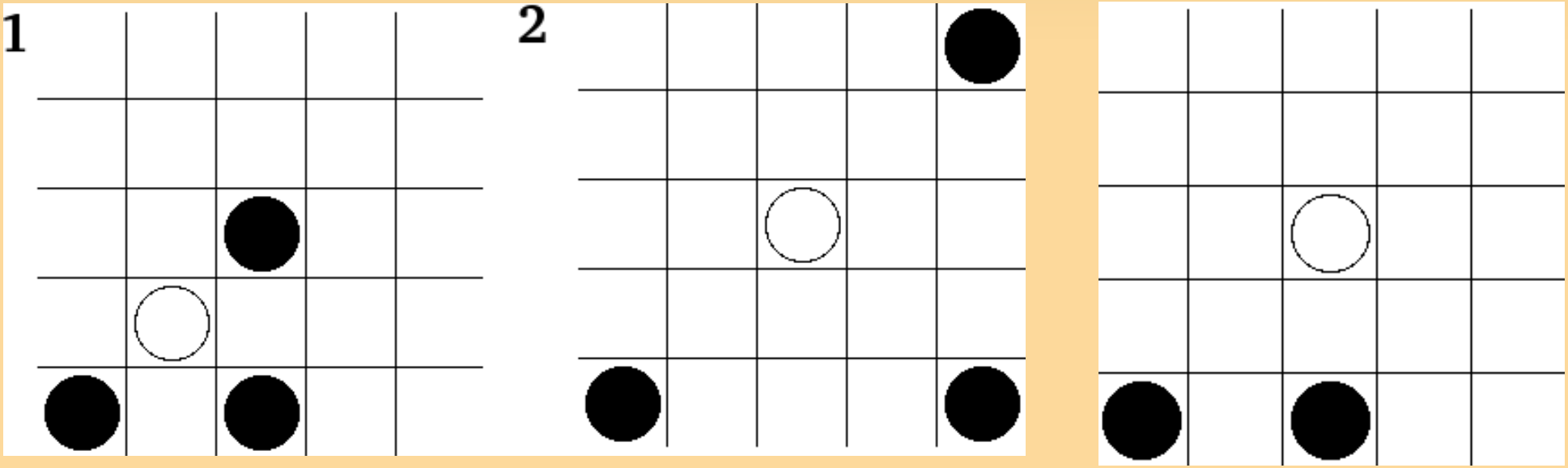
3.2 Optimisation préliminaire

L'espace étant torique, les états sont similaires par translation.



4.1 Expérimentation

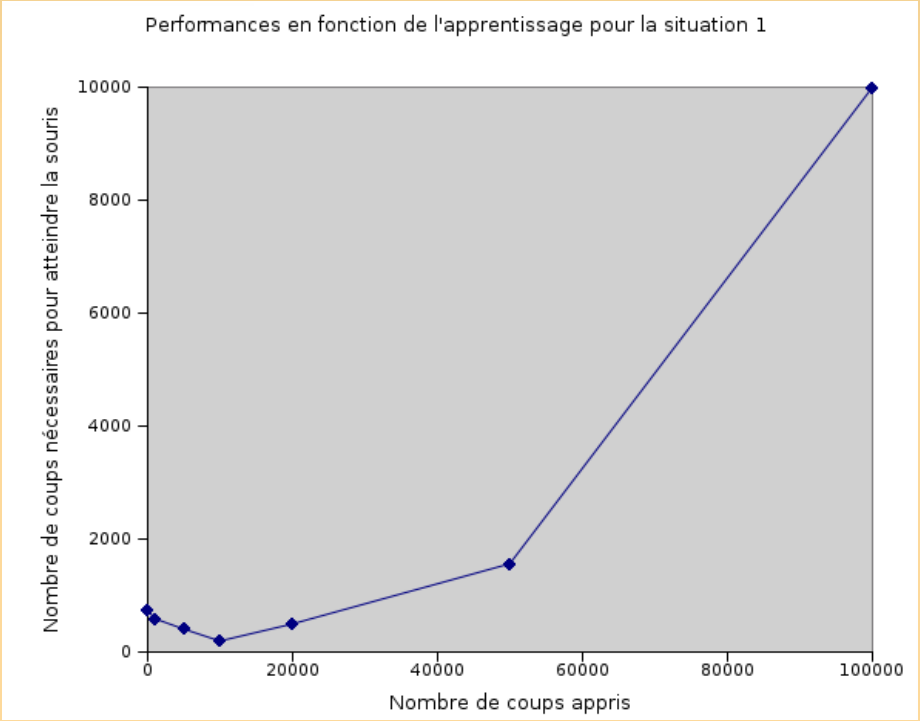
3 chats et une souris - 2 chats et une souris fatiguée

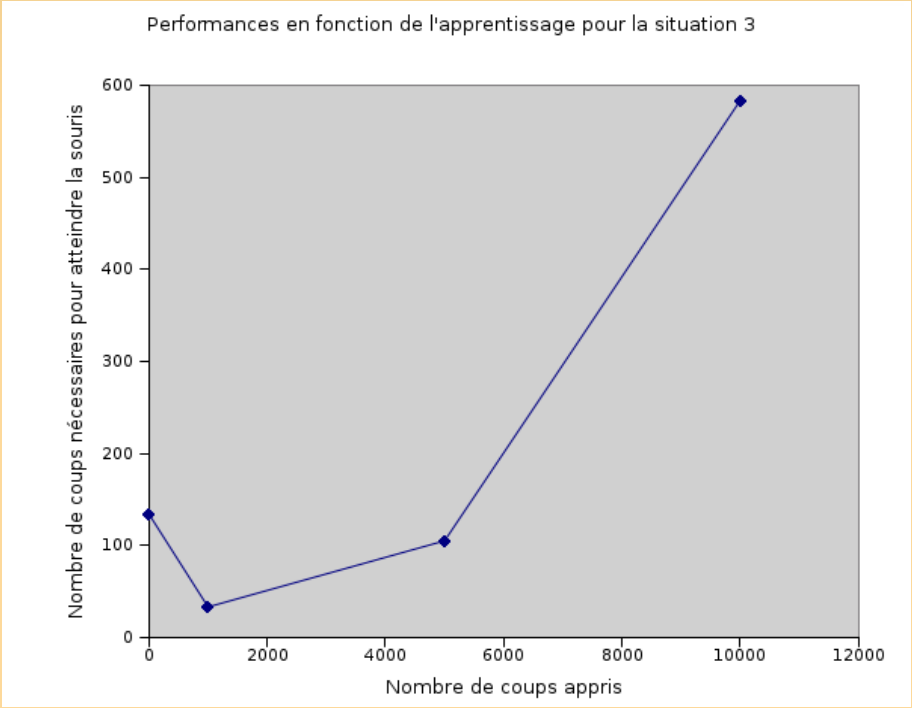


15 min

4.2 Résultats

Fort overfitting avec le nombre de coups d'apprentissage.





4.3 Perspectives

- Performances décevantes -> mieux comprendre l'overfitting, étudier les cas où l'aléatoire n'a pas la tâche aussi facile
- Grande complexité spatiale de Q -> développer une notion de similarité entre les états
- Affiner la coopération en individualisant le reward aux sous-agents

Questions ?

Des questions ?

Bibliographie :

[Littman 1994] Michael L. Littman, *Markov games as a framework for multi-agent reinforcement learning*, In Proceedings of the 11th International Conference on Machine Learning (ML-94), 1994

[Achbany et al, 2006] Youssef Achbany, François Fouss, Luh Yen, Alain Pirotte & Marco Saerens, *Managing the Exploration/Exploitation Trade-Off in Reinforcement Learning*, submitted for publication, 2006