

Q-Learning



Le génie pour l'industrie

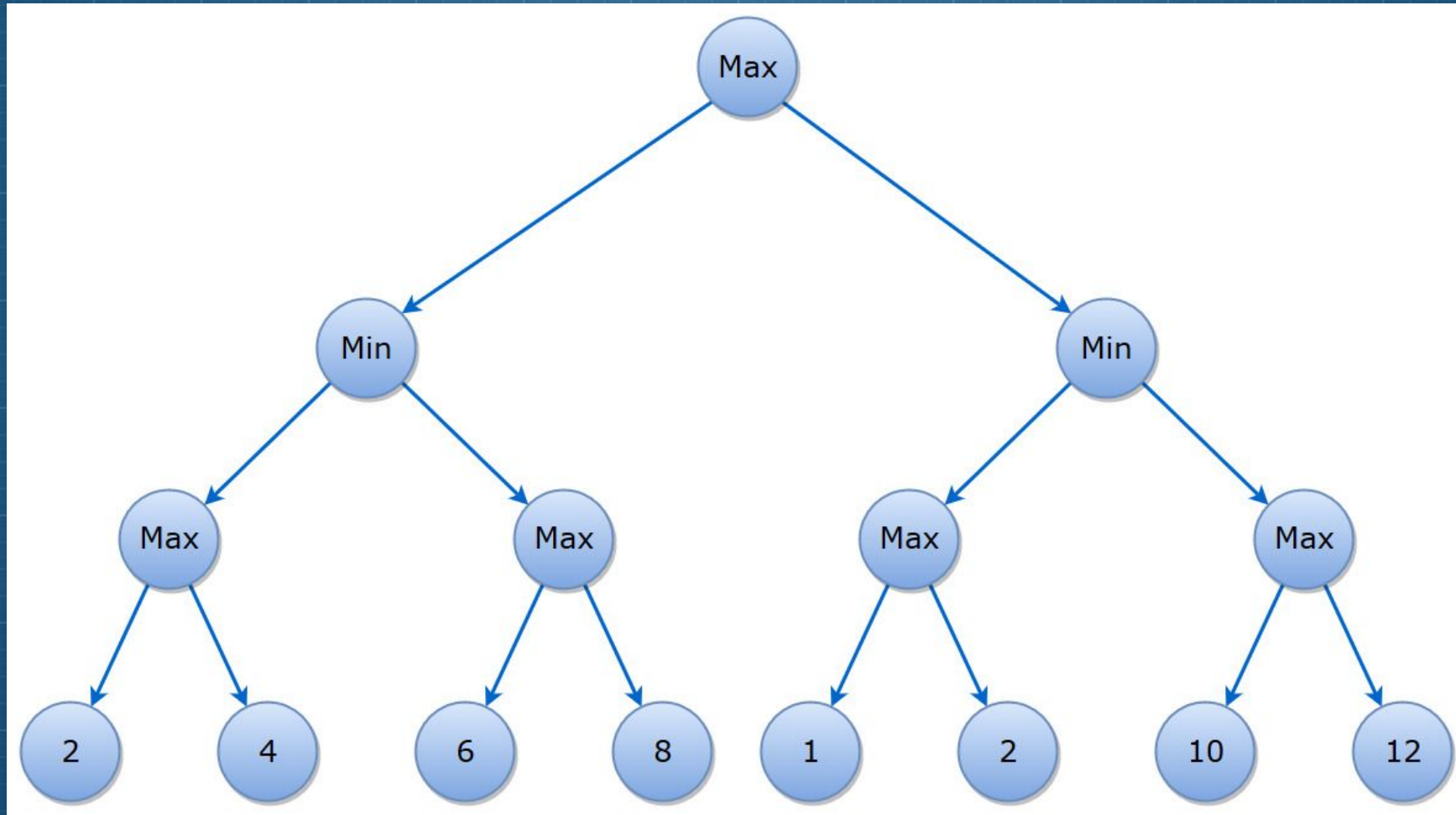
par
Thierry Pouplier



Sujets

1. Min-Max
2. Markov Decision Process
3. Discounting
4. Q-Learning
5. Exploration
6. Q-learning Approximé

Min-Max



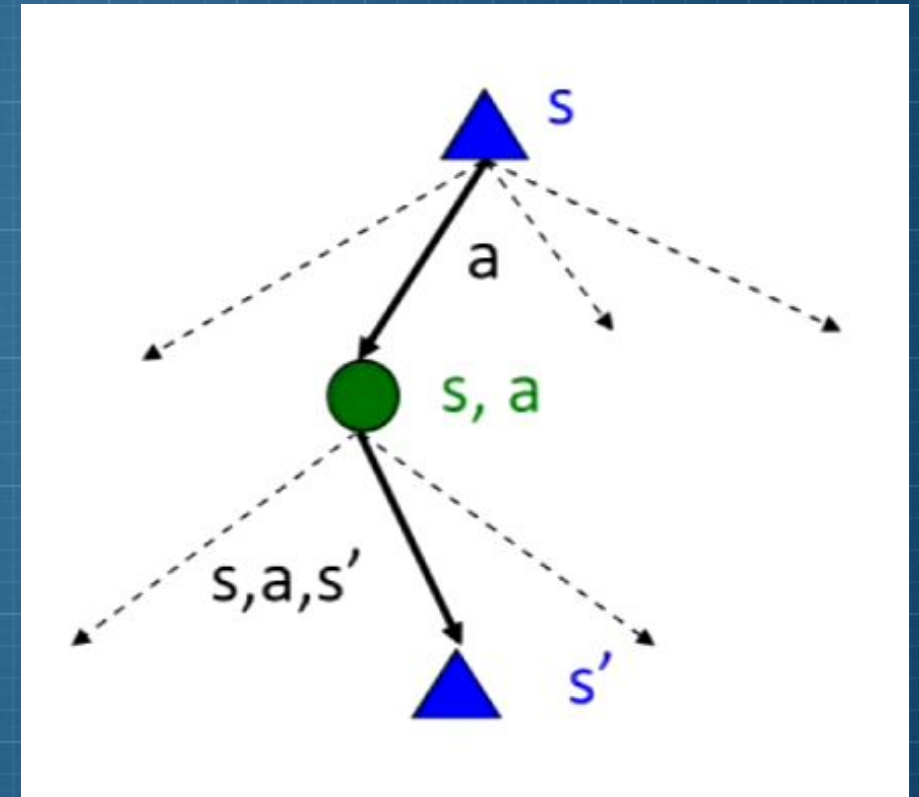
Markov Decision Processes

1. Markov decision processes

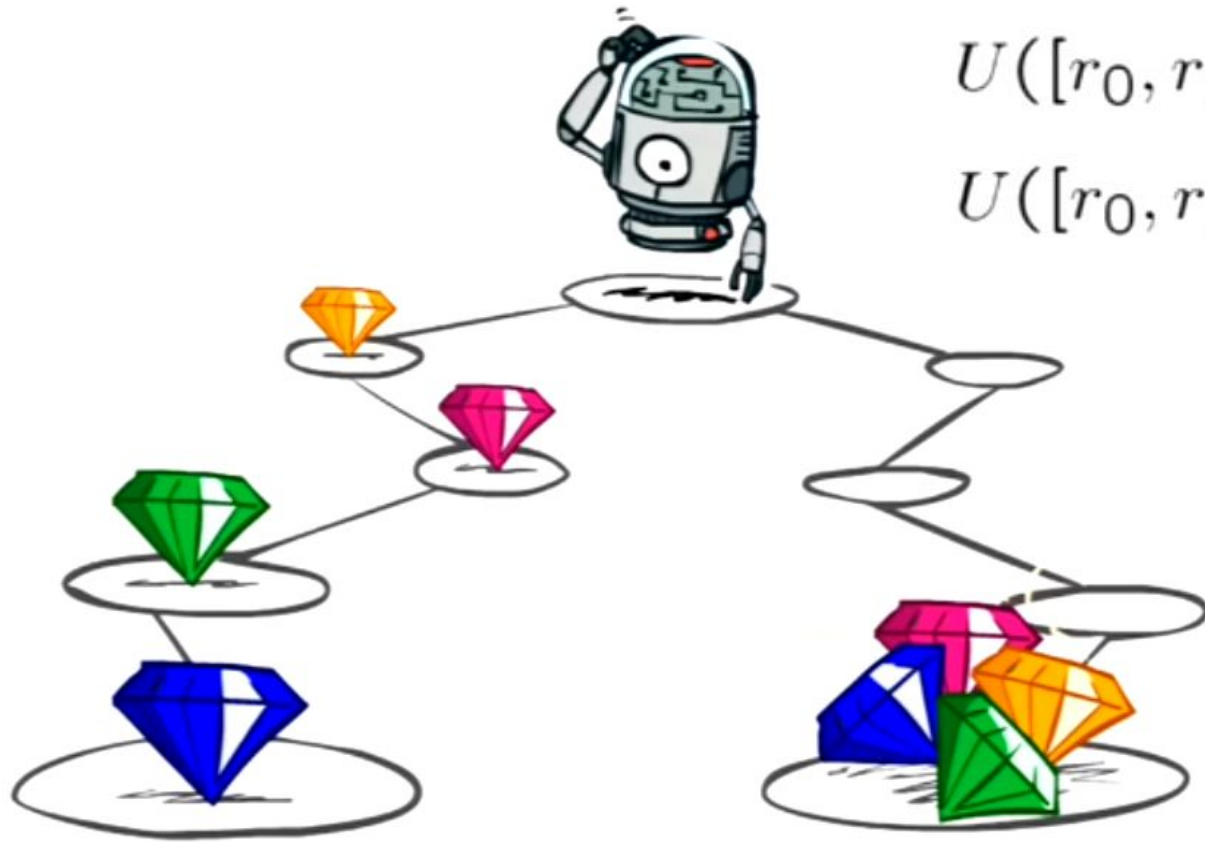
- States S
- Action A
- Transition $T(s,a,s')$ ou $P(s' | s,a)$
- Rewards $R(s,a,s')$

2. Quantités

- Policy = map of states to actions
- Utility = sum of discounted rewards
- Value = expected future utility from a state (max node)
- Q-Value = expected future utility from a q-state (chance node)



Discounting



$$U([r_0, r_1, r_2, \dots]) = r_0 + r_1 + r_2 + \dots$$

$$U([r_0, r_1, r_2, \dots]) = r_0 + \gamma r_1 + \gamma^2 r_2 \dots$$

Q-Learning

$$Q_{k+1}(s, a) \leftarrow \sum_{s'} T(s, a, s') [R(s, a, s') + \gamma \max_{a'} (Q_k(s', a'))]$$

$$Q(s, a) \approx r + \gamma \max_{a'} (Q(s', a'))$$

$$Q(s, a) \leftarrow (1 - \alpha) Q(s, a) + (\alpha) [r + \gamma \max_{a'} (Q(s', a'))]$$

Exploration

- ϵ -greedy

Petite probabilité ϵ de faire n'importe quoi

Grande probabilité $(1-\epsilon)$ de suivre la policy apprise jusqu'à présent

- Exploration function

$$u + k/n$$

Q-Learning Approximé

Problème avec le Q-learning → Beaucoup trop gros tableau de Q-Value !!
(Parce qu'il y a trop d'états!)

Autre problème →



Q-Learning Approximé

$$Q(s, a) = w_1 f_1(s, a) + w_2 f_2(s, a) + \dots + w_n f_n(s, a)$$

Update des poids:

$$\text{difference} = \left[r + \gamma \max_{a'} Q(s', a') \right] - Q(s, a)$$

$$Q(s, a) \leftarrow Q(s, a) + \alpha [\text{difference}]$$

$$w_i \leftarrow w_i + \alpha [\text{difference}] f_i(s, a)$$

Les défis ! ;P

Download:

<http://ai.berkeley.edu/projects/release/reinforcement/v1/001/reinforcement.zip>

Défis:

<http://ai.berkeley.edu/reinforcement.html#Q4>

<http://ai.berkeley.edu/reinforcement.html#Q5>

<http://ai.berkeley.edu/reinforcement.html#Q6>

<http://ai.berkeley.edu/reinforcement.html#Q7>

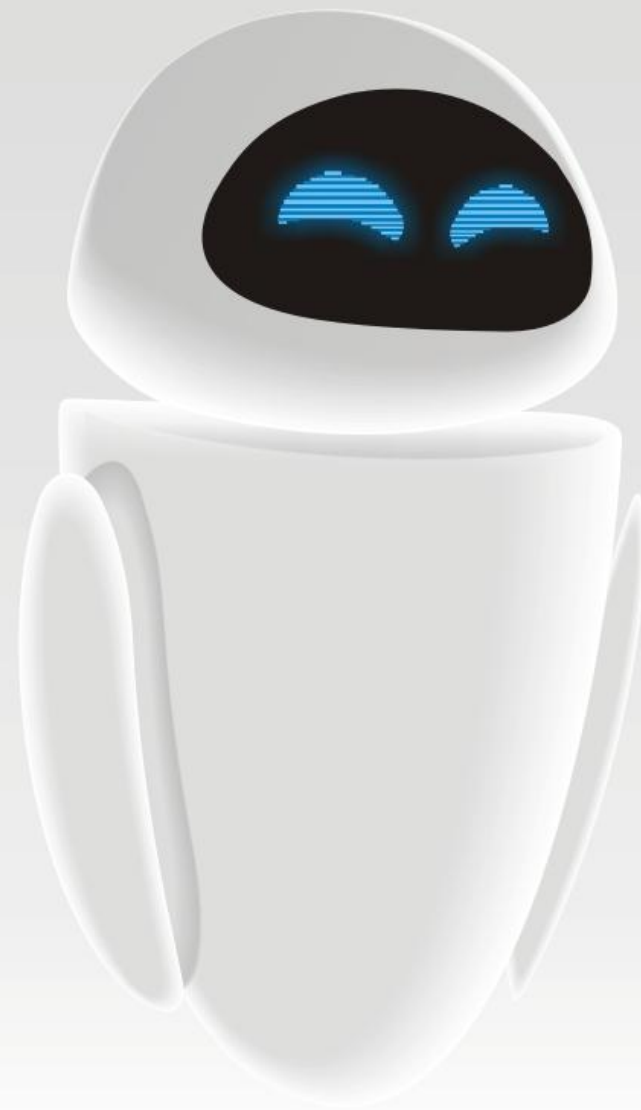
<http://ai.berkeley.edu/reinforcement.html#Q8>

Playlist de la vie:

Playlist: [COMPSCI 188](#)

Plus particulièrement les vidéos suivant:

- [COMPSCI 188 - 2018-09-18](#)
- [COMPSCI 188 - 2018-09-20](#)
- [COMPSCI 188 - 2018-09-25](#)
- [COMPSCI 188 - 2018-09-27](#)



Questions ?