Janvier 2021, Paris

Processus de Décisions Markovien

Calic Petar, Muyang Shi

Encadré par Emmanuel Hyon

Sorbonne Université / Master ANDROIDE

# SOMMAIRE

### Titre - page 1

### Sommaire - page 2

### Introduction - page 3

### Titre - page x

### Titre - page x

### Bibliographie - page x

# Introduction

En théorie des probabilités, un **processus de décision markovien** (MDP) est un modèle stochastique ( de fonction aléatoire) où un agent prend des décisions et où les résultats de ses actions sont aléatoires. Les MDPs sont une extension des chaînes de Markov avec plusieurs actions à choisir par état et où des récompenses sont gagnées par l'agent.

* On y fait une supposition markovienne (de premier ordre) sur la distribution des états visités
* Requière qu’on décrive un objectif à atteindre à partir d’une fonction de récompense basée seulement sur l’état courant

Les MDPs sont utilisés pour étudier des problèmes d'optimisation à l'aide d'algorithmes de programmation dynamique ou d'apprentissage par renforcement, que nous allons étudier et comparer.

## Définition formelle

Un MDP est un quadruplet **{S, A,T, R}** définissant :

* un *ensemble d'états* *S*, qui peut être fini, dénombrable ou continu; cet ensemble définit l'environnement tel que perçu par l'agent (dans le cas d'un robot, on peut voir cela comme l'ensemble produit des valeurs de ses différents capteurs);
* un *ensemble d'actions* *A*, qui peut être fini, dénombrable ou continu et dans lequel l'agent choisit les interactions qu'il effectue avec l'environnement (dans le cas d'un robot on peut voir cela comme l'ensemble produit des paramètres de ses différentes commandes);
* une *fonction de transition* *T : S x A x S 🡪 [0;1];* cette fonction définit l'effet des actions de l'agent sur l'environnement: *T(s, a, s’)* représente la probabilité de se retrouver dans l'état *s'* en effectuant l'action *a*, sachant que l'on était à l'instant d'avant dans l'état *s*.
* une *fonction de récompense R : S x A x S x |R 🡪 [0;1];* elle définit la récompense (positive ou négative) reçue par l'agent: *R(s,a,s',v)* est la probabilité d'obtenir une récompense *v* pour être passé de l'état *s* à *s'* en ayant effectué l'action *a*. Ici encore cette définition est très générale, bien souvent on se contentera par exemple des cas particuliers suivants :
  + - *R : S x A x S 🡪 |R* (récompense déterministe, c'est le choix que nous adopterons dans la suite) ;
    - *R : S x A 🡪 |R* (récompense déterministe rattachée à l'action en ignorant son résultat) ;
    - *R : S 🡪 |R* (récompense déterministe rattachée à un état donné).

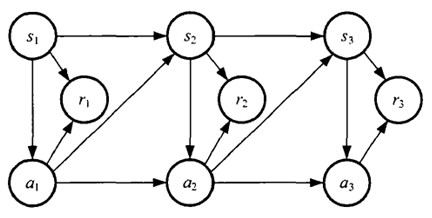
## Processus dynamique du MDP

Soit l’état initial d’un agent *s0*, on choisit une action *a0* dans *A* à exécuter. Après la exécution, l’agent a transféré aléatoirement à l’état prochain *s1* par rapport à la proba de *T, s1∈T0a0*. Et puis l’action *a1*, passer à *s2*, et puis *a2*… On peut représenter cette processus comme la figure ci-dessous :

文本

中度可信度描述已自动生成

Si la récompense a obtenu par rapport aux états *s* et aux actions *a*, le MDP peut être représenté comme la figure ci-dessous :



# Bibliographie

Wikipedia : <https://fr.wikipedia.org/wiki/Processus_de_d%C3%A9cision_markovien>

Emmanuel Hyon : Chapter 1, An introduction to Markov Decision

Processes

Frédérick GARCIA : Chapitre 1, Processus Décisionnels de Markov