

Cahier des charges: Comparaison de solveurs pour MDP

Janvier 2021

Introduction

Nous avons pour objectif dans ce projet, de faire une comparaison numérique de solveurs pour MDP. Afin de faire cela, nous allons d'abord faire des recherches sur les fondements des MDP pour les maîtriser. Ceci est essentiel puisque nous devrons ensuite prendre en main chacun des 3 solveurs que nous allons étudier. Nous allons ensuite préparer un benchmark qui doit être appliqué à tous les serveurs. La dernière phase est celle où on doit faire tourner les solveurs sur nos exemples et analyser les résultats obtenus.

Les Processus de décisions Markoviens

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
- Requiert 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 stochastique. Ils sont résolus par des algorithmes de programmation dynamique que nous allons étudier et comparer.

Definition Formelle:

Un MDP est un quadruplet $\{ \mathbf{S}, \mathbf{A}, \mathbf{T}, \mathbf{R} \}$ où:

- \mathbf{S} , un ensemble d'état représentant l'environnement. Il peut être fini ou continu.
- \mathbf{A} , un ensemble d'action, représentant les actions qui peuvent être prises à un état donnée et mènent vers un autre ou même état de \mathbf{S} . Il est aussi fini ou continu.
- \mathbf{T} , la fonction de transition qui associe à un triplet $(\mathbf{s}, \mathbf{a}, \mathbf{s}')$ la probabilité de se trouver à l'état \mathbf{s}' en effectuant l'action \mathbf{a} à l'état \mathbf{s} . Dans un MDP déterministe un doublet (\mathbf{s}, \mathbf{a}) suffit: \mathbf{A} chaque fois on sera envoyé dans le même état d'arrivée.
- \mathbf{R} , une fonction de récompense: elle définit la récompense (positive ou négative) reçue par l'agent suite à une action de \mathbf{A} .

Poser un problème décisionnel de Markov, c'est trouver une politique (pour chaque état déterminer l'action de telle manière à ce que le gain soit le plus grand à long terme) qui optimise un critère de performance qu'on a établi auparavant comme par exemple le critère fini (Somme de toutes les récompenses accumulées) et le critère pondéré (critère à horizon infini classique, puisqu'on s'arrange que les coefficients tendent vers 0 avec le temps).

Objectifs du Projet

Maitriser les MDP

Avant de pouvoir manipuler les solveurs, il est crucial de comprendre les fondements théoriques des MDP. Tout d'abord il faut savoir reconnaître les situations dans lesquelles un problème peut être modélisé par un MDP. Des hypothèses telles que l'hypothèse de Markov doivent être vérifiées. Ensuite il va falloir savoir modéliser un problème sous forme de MDP, puis adapter le modèle au solveur qu'on souhaite utiliser.

Prendre en main les Solveurs

L'objectif principal de ce projet est de comparer 3 solveurs de MDP. Nous allons observer les performances de Gurobi (programme mathématique) MarmotMDP et ToolboxMDP. Avant de faire cela il va falloir bien s'approprier les 3 solveurs et savoir les manipuler. Trois grandes familles de méthodes existent pour résoudre de tels MDP en cherchant la politique optimale:

- Value Iteration: L'approche la plus classique qui se base aussi sur la résolution directe de l'équation d'optimalité de Bellman \mathbf{V} . On fait des itérations sur les états jusqu'à convergence.

-
- Policy Iteration: Iteration sur la politique: On propose une politique de départ et chaque itération fait une évaluation puis amélioration de la politique jusqu'à convergence.
 - Programmation linéaire. Transformation de l'MDP en un problème linéaire et son Dual pour le résoudre.

Selon les critères, les algos ne sont pas les mêmes à utiliser. Les types de résultats(politiques) ne sont pas les mêmes aussi.

Modéliser un problème MDP pour solveur

Pour pouvoir comparer les solveurs, nous allons prendre un exemple de problème modélisable comme un MDP et le résoudre. Il y a une étape intermédiaire importante et c'est la modélisation du problème et donc son adaptation ou traduction pour un solveur spécifique. En effet, chaque solveur requiert une modélisation particulière d'un problème pour pouvoir le résoudre.

Analyse des solveurs

C'est l'étape finale et clé du projet. Nous allons faire tourner les 3 solveurs sur un même problème et observer les résultats. Nous allons construire un benchmark (Un ensemble d'exemples pour tester les caractéristiques) afin de comparer leurs caractéristiques comme la vitesse d'exécution, l'optimalité de la solution, la précision, la complexité d'utilisation.

Nous allons ensuite faire une synthèse des caractéristiques de chaque solveur et noter leurs avantages et inconvénients. Nous allons déterminer aussi pour chaque type de problème, quel solveur serait le plus adapté et performant.

Planning(estimation):

- Recherche et maîtrise des MDP: 01 février - 30 février
- Prise en main des solveurs: 01 mars - 21 mars
- Préparation du benchmark: 21 mars - 14 avril
- Analyse des résultats: 14 avril - 25 avril
- Rédaction du rapport : 25 avril - 10 mai

Cela correspond à 3 mois et demi de travail. Ceci est notre estimation de temps nécessaire pour faire un travail de qualité sans grande pression du temps.

L'ensemble du projet sera coder sous Python à l'aide d'une machine virtuelle sous Linux. Le rapport sera écrit à l'aide de Overleaf en Latex.