Intelligence Artificielle

Planification stochastique & MDP

5.1 BridgeGrid

a. Changez un seul des deux paramètres, soit gamma soir le bruit, de sorte à ce que la politique optimale permette à l'agent de traverser le pont.

Dans cet environnement, on peut remarquer que les états négatifs ont une valeur très forte (-100) comparée à celle de l'état d'arrivée (10). Ainsi en prenant en compte les paramètres initiaux, l'agent ne prendra pas le risque de traverser le pont puisqu'il a une probabilité de dévier de sa trajectoire et donc d'aller dans un état négatif.

En prenant en compte ces observations, il suffit de réduire le **bruit à zéro.** De cette manière, l'agent sera complètement déterministe et ne pourra donc pas dévier de sa trajectoire. Ainsi, l'agent à tout intérêt à aller à droite étant donné que les valeurs de chaque case seront relativement proche de 10, indiquant une récompense proche.

Figure 1 - Politique permettant à l'agent de traverser le pont

5.2 DiscountGrid

a. Qui suit un chemin risqué pour atteindre l'état absorbant de récompense +1

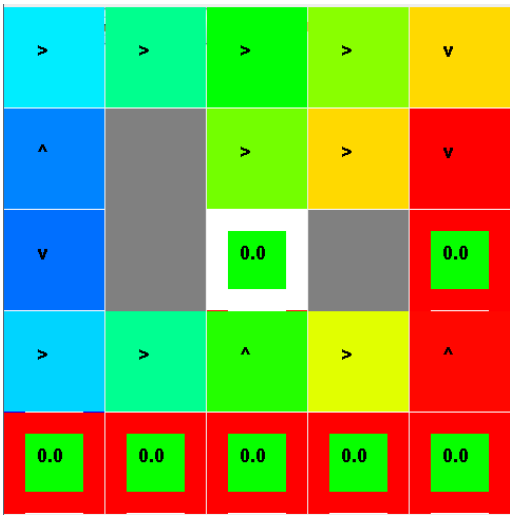

Figure 2 - Politique optimale pour atteindre l'état +1 par le chemin risqué

Dans la configuration initiale, nous obtenons une politique optimale pour atteindre l'état +10 avec le chemin risqué.

Ainsi, pour rester dans le chemin risqué tout en favorisant l'état +1, il faut jouer sur la récompense donnée sur les états normaux. En mettant une **récompense négative de -2**, l'agent n'a pas intérêt à emprunter le chemin sûr qui ne lui promet pas de récompense à court terme.

Cette récompense négative permet également à l'agent de favoriser l'état +1 car il s'agit du premier état positif sur le chemin risqué : il préférera donc l'emprunter plutôt que de prendre le risque d'aller sur un autre état négatif, même s'il amène sur l'état +10.

b. Qui suit un chemin risqué pour atteindre l'état absorbant de récompense +10

En suivant la logique de la configuration BridgeGrid, il suffit de mettre le **bruit à zéro** pour que l'agent prenne le chemin risqué (et le plus court).

Figure 3 - Politique permettant d'atteindre l'état +10 en suivant un chemin risqué

c. Qui suit un chemin sûr pour atteindre l'état absorbant de récompense +1

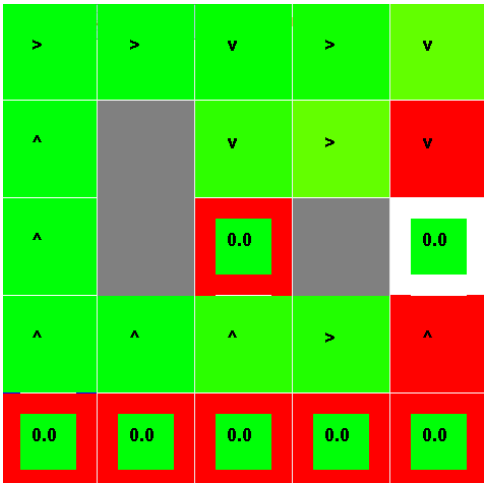

Figure 4 - Politique optimale pour atteindre l'état +1 par le chemin sûr

Ici, nous remarquons au vu des manipulations précédentes qu'influer sur le bruit ne contribuera pas à faire passer l'agent par le chemin sûr : en effet, s'il n'a aucune probabilité de dévier, il prendra forcément le chemin le plus court, qui est le chemin risqué.

De plus, donner des récompenses négatives bloque le chemin sûr comportant plus d'états normaux pour arriver à l'état +1.

Il ne reste donc plus qu'à influer sur gamma. Pour cela, on va diminuer ce paramètre, ce qui va diminuer « l'horizon des récompenses » de l'agent : celui-ci ne verra pas immédiatement l'état +10 en empruntant le chemin sûr et préfèrera donc aller sur l'état +1 (**gamma = 0.25).**

d. Qui évite les états absorbants

Pour éviter les états absorbants, on souhaite que l'agent ne se déplace que sur les états normaux.

Ainsi, en donnant une **récompense (=11)** supérieure à tous les états absorbants aux états normaux, l'agent va éviter les états absorbants qui n'ont plus d'intérêt pour lui étant donné que les états normaux lui procurent plus de récompenses.

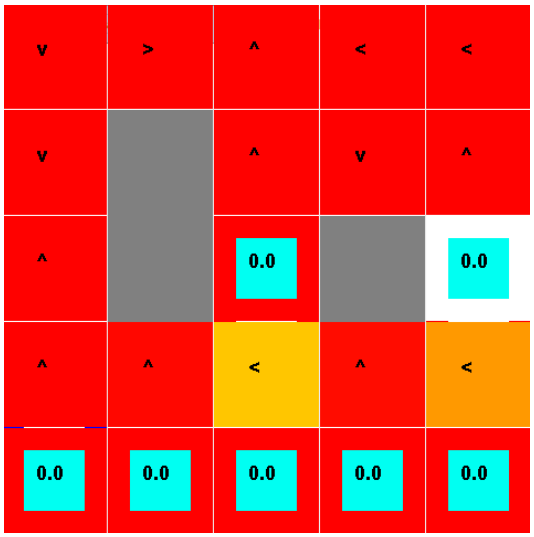

Figure 5 - Politique optimale évitant les états absorbants