

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Iter | Rain | Cloudy | Sunny | Meteor |
| 0 | 0 | 0 | 0 | 0 |
| 1 | -1,8 | 0,5 | 0,8 | 0 |
| 2 | -3,37 | 0,63 | 1,43 | 0 |

1. Wanneer de discount value 0 is betekent het dat het algoritme dat de beslissingen maakt niets geeft om de waarde van de rewards in de toekomst. Wanneer de discount value 1 is heeft dat weer het omgekeerde effect. De discount value geeft dus steeds meer om de toekomst des te hoger de waarde, maar het is bijna nooit het geval dat een actie zijn consequenties oneindig ver in de toekomst er net zo veel toe doet. Bij een discount value van 1 zal er dus veel te veel irrelevante data worden meegenomen bij hem maken van een beslissing. Dit is te zien met de formule  
 *Gt*=∑∞*k*=0 *γ^k\*Rt*+*k*+1 , waar te zien is dat bij een discount value van 1 de Reward van k = 1 net zo belangrijk zal zijn als de Reward van k = 2

2. Volgens 'Markov games as a framework for multi-agent reinforcement learning by Michael Littman, 1994' Kan je de discount value zien als de kans wanneer de simulatie eindigt. Wanneer de discount value 1 is zeg je, in theorie, dat de simulatie oneindig door zal gaan. In de praktijk is dat natuurlijk nooit het geval. In het echt zal er altijd een einde komen aan de simulatie.

|  |  |  |  |
| --- | --- | --- | --- |
| Iter | S1 (linker) | S2 (middle) | End State |
| 0 | 0 | 0 | 0 |
| 1 | -0,1 | -0,1 | 0 |
| 2 | -0,2 | -0,2 | 0 |
| 3 | -0,3 | -0,3 | 0 |
| 4 | -0,4 | -0,4 | 0 |
| 5 | -0,5 | -0,5 | 0 |
| 6 | -0,6 | -0,6 | 0 |
| 7 | -0,7 | -0,7 | 0 |
| 8 | -0,8 | -0,8 | 0 |
| 9 | -0,9 | -0,9 | 0 |
| 10 | -1 | -1 | 0 |
| 11 | -1,1 | -1 | 0 |
| 12 | -1,1 | -1 | 0 |
| 13 | -1,1 | -1 | 0 |
| 14 | -1,1 | -1 | 0 |

S2 value, iter 1 = |(-0,1 + 1 \* 0 = -0,1) ; (-1 + 1 \* 0 = -1)| = **-0,1**  
S1 value, iter 1 = -0,1 + 1 \* 0 = **-0,1**

S2 value, iter 2 = |(-0,1 + 1 \* -0,1 = -0,2) ; (-1 + 1 \* 0 = -1)| = **-0,2**  
S1 value, iter 2 = -0,1 + 1 \* -0,1 = **-0,2**

S2 value, iter 10 = |(-0,1 + 1 \* -0,9 = -1) ; (-1 + 1 \* 0 = -1)| = **-1**  
S1 value, iter 10 = -0,1 + 1 \* -0,9 = **-1**

S2 value, iter 11 = |(-0,1 + 1 \* -1 = -1,1) ; (-1 + 1 \* 0 = -1)| = **-1**  
S1 value, iter 11 = -0,1 + 1 \* -1 = **-1,1**

S2 value, iter 12 = |(-0,1 + 1 \* -1,1 = -1,2) ; (-1 + 1 \* 0 = -1)| = **-1**  
S1 value, iter 12 = -0,1 + 1 \* -1 = **-1,1**

We zien hier dat vanaf iteratie 11 de waardes niet meer veranderen. Gezien het dan geen nut meer heeft om dan verder te gaan itereren kunnen we daar stoppen.

De waardes stoppen met veranderen omdat S2 vanaf iteratie 10 niet verder omlaag kan, want de beste actie wordt nu naar de end state gaan in plaats van daar S1 gaan, zoals het heeft gedaan bij de vorige 10 iteraties. Omdat S2 niet meer verandert vanaf iteratie 10, en S1 alleen naar S2 kan gaan, verandert de value van S1 ook niet meer nadat S2 niet meer verandert.