Adaptive system

Model-based prediction control - theorievragen

Naam : Hussin Almoustafa

Studentnummer : 1776495

April 20, 2023

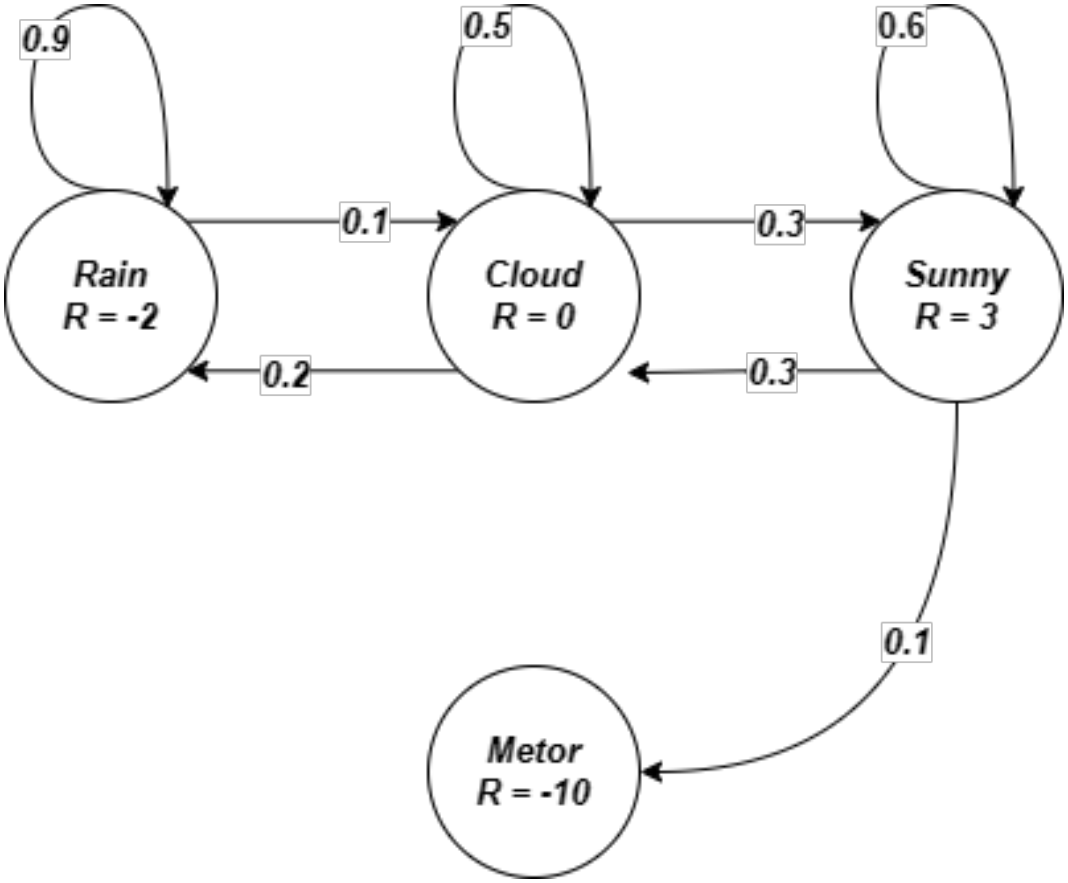

Figure 1: Markov Chain bestaat uit 4 states met reward

1 Opdracht 1 2 - Markov Chain with reward Process

Figure[1] Hierboven Sorry ik heb een latex code error :(

2 Opdracht 3 Sampeling

Markov Chain	rain	cloudy	sunny	meteor
rain	0.9	0.1	0	0
cloudy	0.2	0.5	0.3	0
sunny	0	0.3	0.6	0.1
meteor	0	0	0	1

Markov Reward	rain	cloudy	sunny	meteor
	-2	0	3	-10

Table 1: Markov Chain en Markov Reward

Sample 1: *rain* → *cloudy* → *sunny* → *meteor*

Reward: $-2 + 0 + 3 + (-10) = -9$

Sample 2: *cloudy* → *cloudy* → *rain* → *rain* → *cloudy* → *sunny* → *sunny* → *sunny* → *meteor*

Reward: $0 + 0 + (-2) + (-2) + 0 + 3 + 3 + 3 + (-10) = -5$

3 Opdracht 4 Value function

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
3	-4.77	0.57	1.847	0
4	-6.036	0.3851	2.0792	0
5	-7.19389	0.10911	2.16305	0
6	-8.26359	-0.235308	2.130563	0
7	-9.2607618	-0.6312031	2.0077454	0

Table 2: Value iteration met $\gamma = 1$

De Bellman Expectation Equation met $\gamma = 1$:

$$v_{k+1}(s) = \sum_{a} \pi(a|s) \left(r(s,a) + \sum_{s'} P(s'|s,a) v_k(s') \right) \tag{1}$$

Twee mogelijke problemen met $\gamma = 1$:
Wanneer $= 1$, kunnen sommige problemen leiden tot oneindige returns, wat het moeilijk maakt om een optimale oplossing te vinden. In dergelijke gevallen kan het gebruik van een discount factor ($\gamma < 1$) helpen om de oneindige returns te beperken en een convergente oplossing te vinden.

Het gebruik van een discount factor ($\gamma < 1$) zorgt ervoor dat toekomstige reward minder zwaar wegen dan onmiddellijke reward. Dit kan leiden tot meer realistische oplossingen, aangezien agenten in veel situaties de voorkeur geven aan onmiddellijke reward boven verre toekomstige rewards. Wanneer $\gamma = 1$, wordt er geen onderscheid gemaakt tussen onmiddellijke en toekomstige rewards, wat kan leiden tot onrealistische oplossingen.

4 Opdracht 5 Value iteration.

Iter	S1	S2	S3	S2 meer waard dan S1
1	0	0	0	FALSE
2	-0.1	-0.1	0	FALSE
3	-0.2	-0.2	0	FALSE
4	-0.3	-0.3	0	FALSE
5	-0.4	-0.4	0	FALSE
6	-0.5	-0.5	0	FALSE
7	-0.6	-0.6	0	FALSE
8	-0.7	-0.7	0	FALSE
9	-0.8	-0.8	0	FALSE
10	-0.9	-0.9	0	FALSE
11	-1	-1	0	FALSE
12	-1.1	-1	0	TRUE
13	-1.1	-1	0	TRUE
14	-1.1	-1	0	TRUE

Table 3: Value iteration met $\gamma = 1$ voor een eenvoudige MDP, inclusief een kolom om aan te geven of S2 meer waard is dan S1

Redenering voor het stoppen na de 13e iteratie: Na de 12e iteratie veranderen de waarden van de states niet meer en blijven ze constant. Dit betekent dat de value function is geconvergeerd naar de optimale waarde. Daarom is het niet nodig om verder te itereren.

——— Zie Excel voor formulas toepassing