

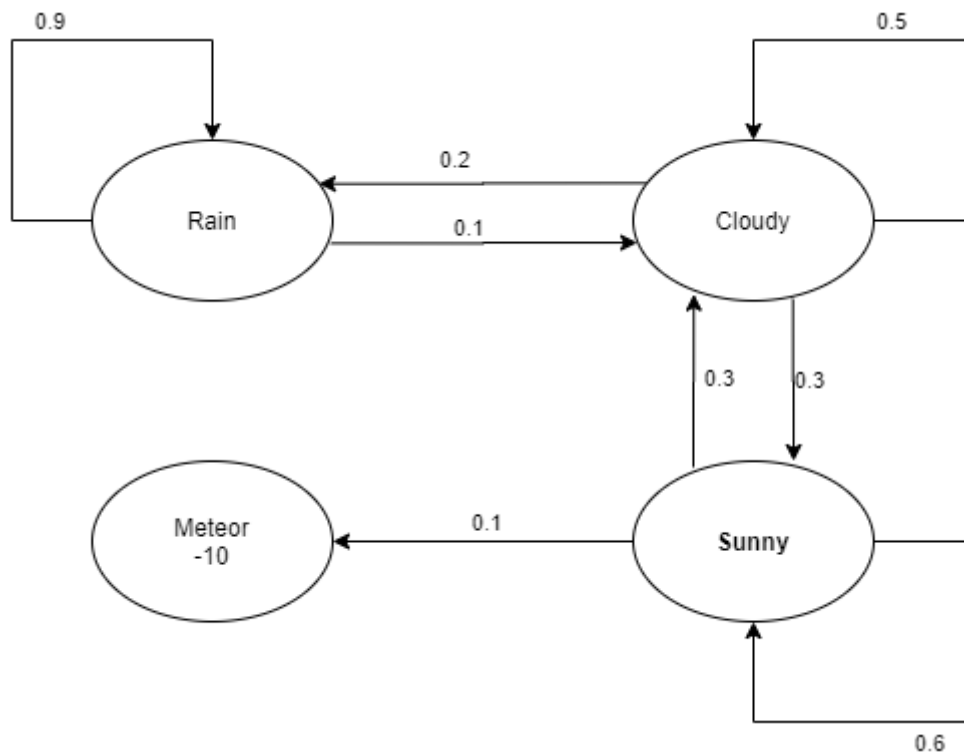


Opdracht 1.1

Adaptive Systems

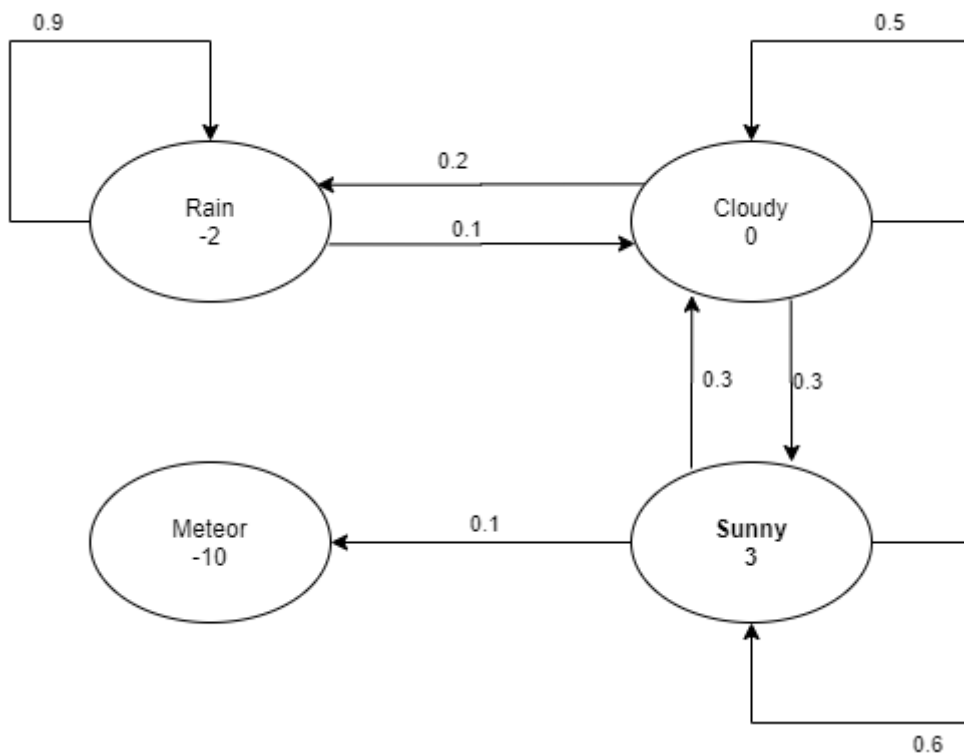
Student:	Storm Joannes
Studentnummer:	1760581
Opleiding:	HBO-ICT Artificial Intelligence
Instelling:	Hogeschool Utrecht
Code:	2022_TICT_VINNO1-33_3_V
Datum:	22-01-2024

A. Markov Chain



...

B. Markov Reward Process



...

C. Sampling

De samples zijn berekend met een discount van 1.

Sample 1: Cloudy → Rainy → Cloudy → Sunny → Meteor
 $G_5: 0 - 2 * 1 + 0 * 1 + 3 * 1 - 10 * 1 = -9$

Sample 2: Rainy → Rainy → Cloudy → Sunny → Sunny → Meteor
 $G_6: -2 - 2 * 1 + 0 * 1 + 3 * 1 + 3 * 1 - 10 * 1 = -8$

D. Value function

Iteratie 1:

$V_r: -2 + 1 * ((0.1 * 0) + (0.9 * 0)) = -2$
 $V_c: 0 + 1 * ((0.2 * 0) + (0.5 * 0) + (0.3 * 0)) = 0$
 $V_s: 3 + 1 * ((0.6 * 0) + (0.3 * 0) + (0.1 * 0)) = 3$
 $V_m: -10$

Iteratie 2:

$V_r: -2 + 1 * ((0.1 * 0) + (0.9 * -2)) = -3.8$
 $V_c: 0 + 1 * ((0.2 * -2) + (0.5 * 0) + (0.3 * 3)) = 0.5$
 $V_s: 3 + 1 * ((0.6 * 3) + (0.3 * 0) + (0.1 * -10)) = 3.8$
 $V_m: -10$

Iteration	Rain	Cloud	Sunny	Meteor
0	0	0	0	0
1	-2	0	3	-10
2	-3.8	0.5	3.8	-10

Discount:

1. Als γ gelijk is aan 1 worden toekomstige beloningen niet verminderd. Dit kan zorgen voor een oneindige beloning.
2. Bij een discount van 1 kijkt hij te veel naar het belang van toekomstige beloningen. Dit kan resulteren in langzame convergentie of instabiliteit.

E. Value iteration

Discount = 1

Beginwaarde van alle states = 0

Iteration 1:

$$A = 1 * (-0.1 + 1 * 0) = -0.1$$

$$B = \text{Max}(0.5 * (-0.1 + 1 * 0), 0.5 * (-1 + 1 * 0)) = -0.05$$

$$C = -1$$

Iteration 2:

$$A = 1 * (-0.1 + 1 * -0.05) = -0.15$$

$$B = \text{Max}(0.5 * (-0.1 + 1 * -0.1), 0.5 * (-1 + 1 * -1)) = -0.1$$

$$C = -1$$

Iteration 3:

$$A = 1 * (-0.1 + 1 * -0.1) = -0.2$$

$$B = \text{Max}(0.5 * (-0.1 + 1 * -0.15), 0.5 * (-1 + 1 * -1)) = -0.125$$

$$C = -1$$

Iteration 4:

$$A = 1 * (-0.1 + 1 * -0.125) = -0.225$$

$$B = \text{Max}(0.5 * (-0.1 + 1 * -0.2), 0.5 * (-1 + 1 * -1)) = -0.15$$

$$C = -1$$

Iteration 5:

$$A = 1 * (-0.1 + 1 * -0.15) = -0.25$$

$$B = \text{Max}(0.5 * (-0.1 + 1 * -0.225), 0.5 * (-1 + 1 * -1)) = -0.1625$$

$$C = -1$$

Iteration 6:

$$A = 1 * (-0.1 + 1 * -0.1625) = -0.2625$$

$$B = \text{Max}(0.5 * (-0.1 + 1 * -0.25), 0.5 * (-1 + 1 * -1)) = -0.175$$

$$C = -1$$

Iteration	A (links)	B (midden)	C (rechts)
0	0	0	0
1	-0.1	-0.05	-1
2	-0.15	-0.1	-1
3	-0.2	-0.125	-1
4	-0.225	-0.15	-1
5	-0.25	-0.1625	-1
6	-0.2625	-0.175	-1

Na 6 iteraties is te zien dat de waarde van state A en B steeds afnemen. Echter is het wel een afnemend dalende lijn. Dit betekent dat het verschil in waarde tussen de iteraties zo klein zal worden dat de waarde van state A nooit lager zal worden dan de waarde van C. Dit zorgt ervoor dat dit MDP met een discount van 1 een oneindige loop is geworden omdat hij nooit voor de end state C zal kiezen.