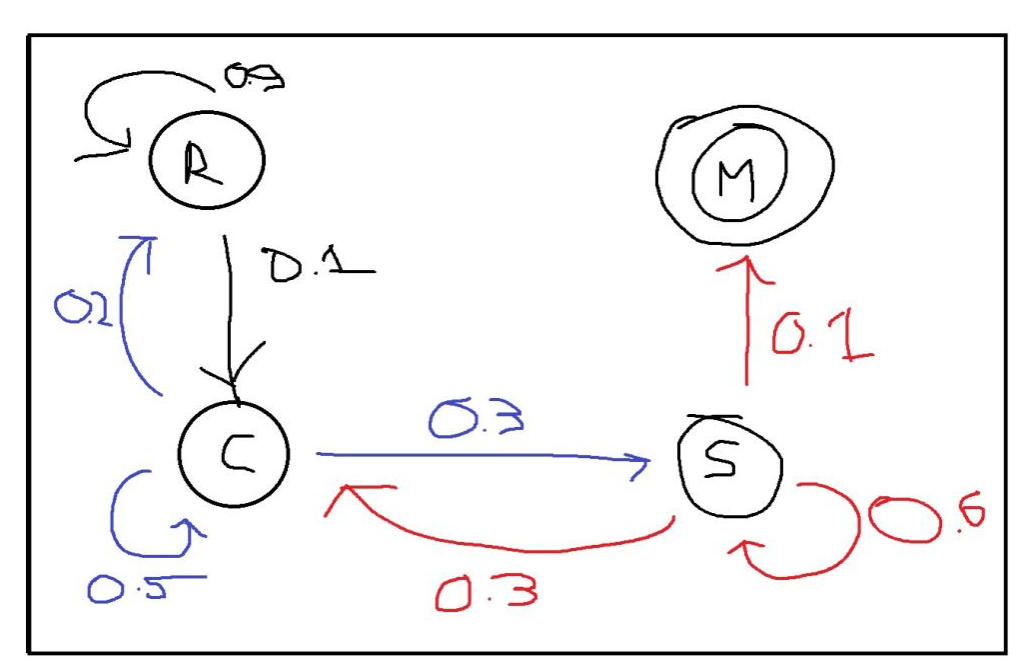
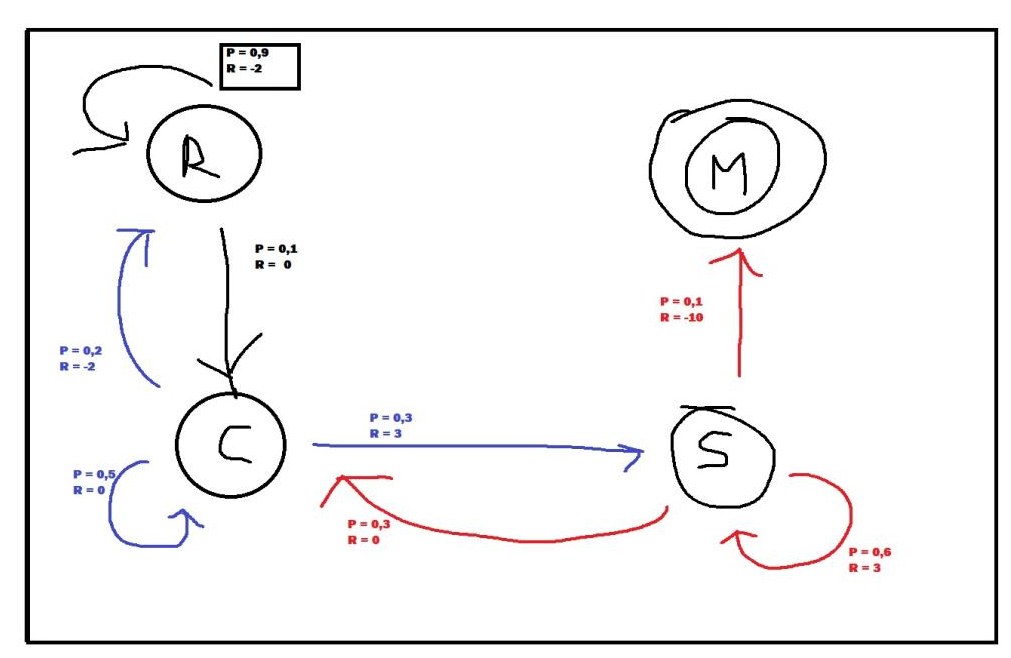
1.1



1.2



1.3

Regen Bewolking Zonnig Zonnig Meteoriet

-2 + (0 x 1) + (3 x 1) + (3 x 1) + (-10 x 1) = -6

Zonnig Regen Bewolking Zonnig Bewolking Zonnig Meteoriet

3 + (-2 x 1) + (0 x 1) + (3 x 1) + (0 x 1) + (3 x 1) + (-10 x 1) = -3

1.4

|  |  |  |  |
| --- | --- | --- | --- |
| State | Value Iteration 0 | Value Iteration 1 | Value Iteration 2 |
| Regen | 0 | -1.8 | -3.37 |
| Bewolkt | 0 | 0.5 | 0.63 |
| Zonnig | 0 | 0.8 | 1.33 |
| Meteoor | 0 | 0 | 0 |

Iteratie 2 uitwerking:

0.6 \* (3 + 0.8) + 0.3 \* (0 + 0.5) + 0.1 \* (-10 + 0)

0.6 \* (3 + 0.8) = 2.28

0.1 \* (0 + 0.5) = 0.05

0.1 \* (-10 + 0) = -1

2.28 + 0.05 + -1 = 1.33

0.5 \* (0 + 0.5) + 0.3 \* (3 + 0.8) + 0.2 \* (-2 + -1.8)

0.5 \* (0 + 0.5) = 0.25

0.3 \* (3 + 0.8) = 1.14

0.2 \* (-2 + -1.8) = -0.76

0.25 + 1.14 + -0.76 = 0.63

0.9 \* (-2 + -1.8) + 0.1 \* (0 + 0.5)

0.9 \* (-2 + -1.8) = -3.42

0.1 \* (0 + 0.5) = 0.05

-3.42 + 0.05 = -3.37

Iteratie 1 uitwerking:

0.6 \* (2 + 0) + 0.3 \* (0 + 0)

----------------------------------

0.6 \* (3 + 0) + 0.3 \* (0 + 0) + 0.1 \* (-10 + 0)

0.6 \* (3 + 0) = 1.8

0.3 \* (0 + 0) = 0

0.1 \* (-10 + 0) = -1

1.8 + 0 + -1 = 0.8

0.5 \* (0 + 0) + 0.3 \* (3 + 0) + 0.2 \* (-2 + 0)

----------------------------------

0.5 \* (0 + 0) = 0

0.3 \* (3 + 0) = 0.9

0.2 \* (-2 + 0) = -0.4

0 + 0.9 + -0.4 = 0.5

0.9 \* (-2 + 0) + 0.1 \* (0 + 0)

----------------------------------

0.9 \* (-2 + 0) = -1.8

0.1 \* (0 + 0) = 0

1.5

Met een discount factor loop je risico dat de values van de states niet convergeren waardoor je dus in een oneindige loop komt te zitten.

Daarnaast bevat je tijd horizon dan te veel rewards voor 1 actie, die eigenlijk niet meer relevant zijn.

2

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| State | Iter 0 | Iter 1 | Iter 2 | Iter 3 | Iter 4 | Iter 5 | Iter 6 | Iter 7 | Iter 8 | Iter 9 | Iter 10 | Iter 11 | Iter 12 |
| 1 | 0.0 | -0.1 | -0.2 | -0.3 | -0.4 | -0.5 | -0.6 | -0.7 | -0.8 | -0.9 | -1 | -1.1 | -1.1 |
| 2 | 0.0 | -0.1 | -0.2 | -0.3 | -0.4 | -0.5 | -0.6 | -0.7 | -0.8 | -0.9 | -1 | -1 | -1 |
| 3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

We stoppen na iteratie 12, omdat de values van state 2 en state 1 niet meer veranderen. De terminal state (state 3) heeft vanaf dit punt altijd de hoogste reward. Het heeft dus geen zin meer om steeds heen en weer te gaan van state 2 naar state 1.

Iteratie 12:

State 2:

-1 + (1 \* 0.0) = -1

-0.1 + (1 \* -1.1) = -1.2

Max(-1, -1.2) = -1

State 1:

-0.1 + (1\*-1) = -1.1

Max(-1.1) = -1.1

Iteratie 11:

State 2:

-1 + (1 \* 0.0) = -1

-0.1 + (1 \* -1) = -1.1

Max(-1, -1.1) = -1

State 1:

-0.1 + (1\*-1) = -1.1

Max(-1.1) = -1.1

Iteratie 10:

State 2:

-1 + (1 \* 0.0) = -1

-0.1 + (1 \* -0.9) = -1

Max(-1, -1) = -1

State 1:

-0.1 + (1\*-0.9) = -1

Max(-1) = -1

Iteratie 9:

State 2:

-1 + (1 \* 0.0) = -1

-0.1 + (1 \* -0.8) = -0.9

Max(-1, -0.9) = -0.9

State 1:

-0.1 + (1\*-0.8) = -0.9

Max(-0.9) = -0.9

Iteratie 8:

State 2:

-1 + (1 \* 0.0) = -1

-0.1 + (1 \* -0.7) = -0.8

Max(-1, -0.8) = -0.8

State 1:

-0.1 + (1\*-0.7) = -0.8

Max(-0.8) = -0.8

Iteratie 7:

State 2:

-1 + (1 \* 0.0) = -1

-0.1 + (1 \* -0.7) = -0.7

Max(-1, -0.7) = -0.7

State 1:

-0.1 + (1\*-0.6) = -0.7

Max(-0.7) = -0.7

Iteratie 6:

State 2:

-1 + (1 \* 0.0) = -1

-0.1 + (1 \* -0.5) = -0.6

Max(-1, -0.6) = -0.6

State 1:

-0.1 + (1\*-0.5) = -0.6

Max(-0.6) = -0.6

Iteratie 5:

State 2:

-1 + (1 \* 0.0) = -1

-0.1 + (1 \* -0.4) = -0.5

Max(-1, -0.5) = -0.5

State 1:

-0.1 + (1\*-0.4) = -0.5

Max(-0.5) = -0.5

Iteratie 4:

State 2:

-1 + (1 \* 0.0) = -1

-0.1 + (1 \* -0.3) = -0.4

Max(-1, -0.4) = -0.4

State 1:

-0.1 + (1\*-0.3) = -0.4

Max(-0.4) = -0.4

Iteratie 3:

State 2:

-1 + (1 \* 0.0) = -1

-0.1 + (1 \* -0.2) = -0.3

Max(-1, -0.3) = -0.3

State 1:

-0.1 + (1\*-0.2) = -0.3

Max(-0.3) = -0.3

Iteratie 2

State 2:

-1 + (1 \* 0.0) = -1

-0.1 + (1 \* -0.1) = -0.2

Max(-1, -0.2) = -0.2

State 1:

-0.1 + (1\*-0.1) = -0.2

Max(-0.2) = -0.2

Iteratie 1

State 2:

-1 + (1 \* 0.0) = -1

-0.1 + (1 \* 0.0) = -0.1

Max(-1, -0.1) = -0.1

State 1:

-0.1 + (1\*0.0) = -0.1

Max(-0.1) = -0.1