Machine Data and Learning

Kushagra Agarwal 2018113012

Assignment 5 Part 1 Report

X: 1- 13/100 = **0.87**		**Y:** 12%3 = **0**

State Space (S):

S={ S1, S2, S3, S4, S5 }

Actions (A):

A={Left, Right}

Transition Probabilities(P):

The probabilities are shown in the form of tables with the initial state in rows and the final states in columns. Value of x used is 0.87 as computed above.

LEFT	s1	s2	s3	s4	s5
s1	0.87	0.13	0	0	0
s2	0.87	0	0.13	0	0
s3	0	0.87	0	0.13	0
s4	0	0	0.87	0	0.13
s5	0	0	0	0.87	0.13

RIGHT	s1	s2	s3	s4	s5
s1	0.13	0.87	0	0	0
s2	0.13	0	0.87	0	0
s3	0	0.13	0	0.87	0
s4	0	0	0.13	0	0.87
s5	0	0	0	0.13	0.87

Set of Observations (Ω) :

Observations : { Red, Green }

Observation Probabilities (O) :

The ending states are in rows and corresponding observations in columns.

LEFT	Red	Green
s1	0.9	0.1
s2	0.9	0.1
s3	0.15	0.85
s4	0.15	0.85
s5	0.9	0.1

RIGHT	Red	Green
s1	0.9	0.1
s2	0.9	0.1
s3	0.15	0.85
s4	0.15	0.85
s5	0.9	0.1

********* To find the belief states we do not need the reward function

Now we know that POMDP solutions with predefined a belief state can be found by using the value iteration algorithm with a modification: Use belief states instead of states. The process is Markovian over belief states and not over states.

Here we have been given the initial belief state b:

The agent can be in any of s1, s2 and s5. Hence the belief state is equally divided between s1,s2 and s5 and has zero probability for s3 and s4.

Initial b = [⅓, ⅓, 0, 0, ⅓] ;

Calculation for b' from b has the following mathematical approach:

s = old state

s' = new state

b = old belief state, and b(s) probability of s given belief state b

a = action

b' = new belief state

b'(s') = probability of s' given b'

o = observation

Ub'(s') = O(s', a, o) * Σ (T(s, a, s') b(s)); summation is over all states

So for example Ub'(s1) = Pr (o | s1, right)) * [Pr(s1 | s1, right) * b(s1) + Pr (s1 | s2, right) * b(s2) + Pr (s1 |s3, right) * b(s3) + Pr (s1 | s4, right) * b(s4) + Pr (s1 | s5, right) * b(s5)]

The U stands for un-normalised, we can calculate Ub'(x) for all other states x, and sum them up to form the denominator. The denominator formula is **Pr (o | a, b),** but we would be using the summation of all Ub' to be the denominator, as we know that summation of all belief states should always be equal to 1

ACTION 1: TOOK RIGHT OBSERVED RED

Initial belief state = **[⅓. ⅓, 0, 0, ⅓]** ; Action **a = Righ**t; Observation **o = Red**

1) State s1

Ub'(s1, right, red) = O(s1, right, red) * Σ (T(s, right , s1) b(s))

Ub'(s1,right) = P(red| s1, right) *[P(s1| s1, right) * b(s1) + P (s1 |s2, right) * b(s2) +P(s1 |s3, right) * b(s3) +P(s1| s4, right) * b(s4) + P (s1| s5, right) * b(s5)]

P(red | s1, right) = 0.9

Ub'(s1, right, red) =0.9* [(0.13* ⅓) + (0.13*⅓) + (0*0) + (0*0) + (0*⅓)]

=0.9*0.26*(⅓) = 0.078

Ub'(s1) =0.078

2) State s2

Ub'(s2, right, red) = O(s2, right, red) * Σ (T(s, right , s2) b(s))

Ub'(s2,right) = P(red| s2, right) *[P(s2| s1, right) * b(s1) + P (s2 |s2, right) * b(s2) +P(s2 |s3, right) * b(s3) +P(s2| s4, right) * b(s4) + P (s2| s5, right) * b(s5)]

P(red | s2, right) = 0.9

Ub'(s2, right, red) =0.9* [(0.87* ⅓) + (0*⅓) + (0.13*0) + (0*0) + (0*⅓)]

=0.9*0.87*(⅓) = 0.261

Ub'(s2) =0.261

3) State s3

Ub'(s3, right, red) = O(s3, right, red) * Σ (T(s, right , s3) b(s))

Ub'(s3,right) = P(red| s3, right) *[P(s3| s1, right) * b(s1) + P (s3 |s2, right) * b(s2) +P(s3 |s3, right) * b(s3) +P(s3| s4, right) * b(s4) + P (s3| s5, right) * b(s5)]

P(red | s3, right) = 0.15

Ub'(s3, right, red) =0.15* [(0* ⅓) + (0.87*⅓) + (0*0) + (0.13*0) + (0*⅓)]

=0.15*0.87*(⅓) = 0.0435

<u>Ub’(s3) =0.0435</u>

4) <u>State s4</u>

Ub’(s4, right, red) = O(s4, right, red) * Σ (T(s, right , s4) b(s))

Ub’(s4,right) = P(red| s4, right) *[P(s4| s1, right) * b(s1) + P (s4 |s2, right) * b(s2) +P(s4 |s3, right) * b(s3) +P(s4| s4, right) * b(s4) + P (s4| s5, right) * b(s5)]

P(red | s4, right) = 0.15

Ub’(s4, right, red) =0.15* [(0* ⅓) + (0*⅓) + (0.87*0) + (0*0) + (0.13*⅓)]

=0.15*0.13*(⅓) = 0.0065

<u>Ub’(s3) =0.0065</u>

5) <u>State s5</u>

Ub’(s5, right, red) = O(s5, right, red) * Σ (T(s, right , s5) b(s))

Ub’(s5,right) = P(red| s5, right) *[P(s5| s1, right) * b(s1) + P (s5 |s2, right) * b(s2) +P(s5 |s3, right) * b(s3) +P(s5| s4, right) * b(s4) + P (s5| s5, right) * b(s5)]

P(red | s2, right) = 0.9

Ub’(s5, right, red) =0.9* [(0* ⅓) + (0*⅓) + (0*0) + (0.87*0) + (0.87*⅓)]

=0.9*0.87*(⅓) = 0.261

<u>Ub’(s2) =0.261</u>

SUM = 0.078 + 0.261 + 0.0435 + 0.0065 + 0.261 = **0.65**

b’=[0.078, 0.261, 0.0435, 0.0065, 0.261] / (0.65)

b’ = [0.12, 0.40153846153, 0.06692307692, 0.01 ,0.40153846153]

ACTION 2: TOOK LEFT OBSERVED GREEN

Initial belief state =**[0.12, 0.40153846153, 0.**06692307692**, 0.01 ,0.40153846153]**

Action **a = Left**; Observation **o=Green**

1) <u>State s1</u>

Ub'(s1,left, green) = O(s1, left, green) * **Σ** (T(s, left , s1) b(s))

Ub'(s1,left) = P(green| s1, left) *[P(s1| s1, left) * b(s1) + P (s1 |s2, left) * b(s2) +P(s1 |s3, left) * b(s3) +P(s1| s4, left) * b(s4) + P (s1| s5, left) * b(s5)]

P(green | s1, left) = 0.1

Ub'(s1, left, green) =0.1* [(0.87* 0.12) + (0.87*0.40153846153) +

(0*0.06692307692) + (0*0.01) + (0*0.40153846153)]

=0.1*0.87*(0.521538461) = 0.045373846

<u>Ub'(s1) =0.045373846</u>

2) <u>State s2</u>

Ub'(s2,left, green) = O(s2, left, green) * **Σ** (T(s, left , s2) b(s))

Ub'(s2,left) = P(green| s2, left) *[P(s2| s1, left) * b(s1) + P (s2 |s2, left) * b(s2) +P(s2 |s3, left) * b(s3) +P(s2| s4, left) * b(s4) + P (s2| s5, left) * b(s5)]

P(green | s2, left) = 0.1

Ub'(s2, left, green) =0.1* [(0.13* 0.12) + (0*0.40153846153) +

(0.87*0.06692307692) + (0*0.01) + (0*0.40153846153)]

= 0.007382307692

<u>Ub'(s2) =0.007382307692</u>

3) <u>State s3</u>

Ub'(s3,left, green) = O(s3, left, green) * **Σ** (T(s, left , s3) b(s))

Ub'(s3,left) = P(green| s3, left) *[P(s3| s1, left) * b(s1) + P (s3 |s2, left) * b(s2) +P(s3 |s3, left) * b(s3) +P(s3| s4, left) * b(s4) + P (s3| s5, left) * b(s5)]

P(green | s3, left) = 0.85

Ub'(s3, left, green) =0.85* [(0* 0.12) + (0.13*0.40153846153) + (0*0.06692307692) + (0.87*0.01) + (0*0.40153846153)]

= 0.051765

<u>Ub'(s3) =0.051765</u>

4) <u>State s4</u>

Ub'(s4,left, green) = O(s4, left, green) * **Σ** (T(s, left , s4) b(s))

Ub'(s4,left) = P(green| s4, left) *[P(s4| s1, left) * b(s1) + P (s4 |s2, left) * b(s2) +P(s4 |s3, left) * b(s3) +P(s4| s4, left) * b(s4) + P (s4| s5, left) * b(s5)]

P(green | s4, left) = 0.85

Ub'(s4, left, green) =0.85* [(0* 0.12) + (0*0.40153846153) + (0.13*0.06692307692) + (0*0.01) + (0.87*0.40153846153)]

= 0.304332692

<u>Ub'(s4) =0.304332692</u>

5) <u>State s5</u>

Ub'(s5,left, green) = O(s5, left, green) * **Σ** (T(s, left , s5) b(s))

Ub'(s5,left) = P(green| s5, left) *[P(s5| s1, left) * b(s1) + P (s5 |s2, left) * b(s2) +P(s5 |s3, left) * b(s3) +P(s5| s4, left) * b(s4) + P (s5| s5, left) * b(s5)]

P(green | s5, left) = 0.1

Ub’(s5, left, green) =0.1* [(0* 0.12) + (0*0.40153846153 + (0*0.06692307692) + (0.13*0.01) + (0.13*0.40153846153)]

= 0.00535

<u>Ub’(s5) =0.00535</u>

SUM = 0.045373846 + 0.007382307692 +0.051765 + 0.304332692 + 0.00535

= **0.414203845**

b’=[0.045373846, 0.007382307692, 0.051765, 0.304332692, 0.00535] / (0.414203845)

b’ = [0.109544724, 0.017822885, 0.124974696, 0.734741349 ,0.012916345]

ACTION 3: TOOK LEFT OBSERVED GREEN

Initial belief state =[0.109544724, 0.017822885, 0.124974696, 0.734741349 ,0.012916345] ;

Action **a = Left**; Observation **o=Green**

1) <u>State s1</u>

Ub’(s1,left, green) = O(s1, left, green) * **Σ** (T(s, left , s1) b(s))

Ub’(s1,left) = P(green| s1, left) *[P(s1| s1, left) * b(s1) + P (s1 |s2, left) * b(s2) +P(s1 |s3, left) * b(s3) +P(s1| s4, left) * b(s4) + P (s1| s5, left) * b(s5)]

P(green | s1, left) = 0.1

Ub’(s1, left, green) =0.1* [(0.87* 0.109544724) + (0.87*0.017822885) + (0*0.124974696) + (0*0.734741349) + (0*0.012916345)]

= 0.011080981

<u>Ub'(s1) =0.011080981</u>

2) <u>State s2</u>

Ub'(s2,left, green) = O(s2, left, green) * **Σ** (T(s, left , s2) b(s))

Ub'(s2,left) = P(green| s2, left) *[P(s2| s1, left) * b(s1) + P (s2 |s2, left) * b(s2) +P(s2 |s3, left) * b(s3) +P(s2| s4, left) * b(s4) + P (s2| s5, left) * b(s5)]

P(green | s2, left) = 0.1

Ub'(s2, left, green) =0.1* [(0.13* 0.109544724) + (0*0.017822885) +

(0.87*0.124974696) + (0*0.734741349) + (0*0.012916345)]

= 0.012296879

<u>Ub'(s2) =0.012296879</u>

3) <u>State s3</u>

Ub'(s3,left, green) = O(s3, left, green) * **Σ** (T(s, left , s3) b(s))

Ub'(s3,left) = P(green| s3, left) *[P(s3| s1, left) * b(s1) + P (s3 |s2, left) * b(s2) +P(s3 |s3, left) * b(s3) +P(s3| s4, left) * b(s4) + P (s3| s5, left) * b(s5)]

P(green | s3, left) = 0.85

Ub'(s3, left, green) =0.85* [(0* 0.109544724) + (0.13*0.017822885) +

(0*0.124974696) + (0.87*0.734741349) + (0*0.012916345)]

= 0.545310656

<u>Ub'(s3) =0.545310656</u>

4) <u>State s4</u>

Ub’(s4,left, green) = O(s4, left, green) * **Σ** (T(s, left , s4) b(s))

Ub’(s4,left) = P(green| s4, left) *[P(s4| s1, left) * b(s1) + P (s4 |s2, left) * b(s2) +P(s4 |s3, left) * b(s3) +P(s4| s4, left) * b(s4) + P (s4| s5, left) * b(s5)]

P(green | s4, left) = 0.85

Ub’(s4, left, green) =0.85* [(0* 0.109544724) + (0*0.017822885) +

(0.13*0.124974696) + (0*0.734741349) + (0.87*0.012916345)]

= 0.023361341

Ub’(s4) =0.023361341

5) State s5

Ub’(s5,left, green) = O(s5, left, green) * **Σ** (T(s, left , s5) b(s))

Ub’(s5,left) = P(green| s5, left) *[P(s5| s1, left) * b(s1) + P (s5 |s2, left) * b(s2) +P(s5 |s3, left) * b(s3) +P(s5| s4, left) * b(s4) + P (s5| s5, left) * b(s5)]

P(green | s5, left) = 0.1

Ub’(s5, left, green) =0.1* [(0* 0.109544724) + (0*0.017822885) + (0*0.124974696) +

(0.13*0.734741349) + (0.13*0.012916345)]

= 0.009719549579

Ub’(s5) =0.009719549579

SUM = 0.011080981 + 0.012296879 +0.545310656 + 0.023361341 + 0.009719549579

= **0.601769406**

b’=[0.011080981, 0.012296879, 0.545310656, 0.023361341, 0.009719549579] / (0.601769406)

b’ = [0.018413998, 0.020434536, 0.906178763, 0.038821084, 0.016151618]